

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N° Réf : .....

Centre Universitaire  
Abd Elhafid Boussouf Mila

Institut des Sciences et Technologie  
Département de Mathématiques et Informatique

## Mémoire préparé en vue de l'obtention du diplôme de Master

En : Informatique  
Spécialité : Sciences et Technologies de l'Information et de la  
Communication (STIC)

### Extraction et Reconnaissance des Entités Nommées à Base d'Ontologie

Préparé par :  
M<sup>lle</sup> BEN MOUHOUB Housna  
M<sup>lle</sup> DERMOUCHI Ryane

Soutenue devant le jury

Président : Mr MEGUEHOUT H.  
Examineur : Mr MRABET A.  
Encadré par : Mr HADJI A.

Grade MCB  
Grade MAA  
Grade MAA

Année Universitaire : 2021/2022

# *Remerciements*

Nous tenons tout d'abord à remercier Dieu le tout puissant Miséricordieux qui nous a donné la force et la patience pour accomplir ce Modest travail.

Un remerciement particulier à nos très chers parents qui nous ont encouragés, soutenu durant tout nos parcours.

En premier lieu, nous exprimons toute notre gratitude pour notre promoteur, Mr Hadji A. pour ses précieux conseils, sa disponibilité, la confiance qu'il nous a toujours témoignée et la sollicitude dont il nous a entouré, et ce tout au long de l'élaboration du présent travail.

Nous tenons à remercier vivement les membres de jury qui ont fait l'honneur d'accepter de participer a la soutenance de ce mémoire.

Nous adressons également nos remerciements, à tous nos enseignants tout au long du cycle d'étude.

À tous ceux qui ont participé de près ou de loin pour l'accomplissement de ce modeste travail.

*Housna @ Ryane*

## *Dédicaces*

**Avec énorme plaisir, un cœur ouvert et une immense joie, que je**  
dédie mon travail à mes très chers, respectueux et magnifique parents  
**m'ont** soutenus tout au long de ma vie

A mes parent « père & mère »

Mes dédicaces vont également à mes frères «Amar, Rachid , Abd El

Alia, Khaled »

Et ma soeur «Imane»

Sans oublier de dédier ce mémoire à mon encadreur Mr Hadji A.

A mon binôme «Housna Ben Mouhoub »

A toutes mes amies «Nada, Chaima, Imane»

En **témoignage de l'amitié sincère qui nous à liée**

Et des bons moments passés ensemble

Je veux remercier tous, vos soutiens et vos encouragements me  
donnent la force de continuer

*Ryane*

## *Dédicaces*

Je dédie mon travail à la nation d'un million et demi de martyrs, mon pays que

**le j'aime trop, l'Algérie**

**Pour la chère à mon cœur, ma mère Nouara, qui était patiente et endurante et**

qui veillait et se fatiguait, sans elle, je n'aurais pas atteint ce point.

A mon héros mon père Abd Allah,

A ma fierté ma seour Moufida

A ma seour Amina et ses adorables filles Insaf et Meriem que Allah les  
protéges.

A la plus solidaire, utile et motivant Dr Nerdjes Bouchemmal

Et tous mes enseignants du département Informatique.

A mon amie et mon Binome Ryane Dermouchi

A mes chères amies Yousra K. et Imane B. que Allah les protéges.

À ma belle amie qui a toujours été avec moi Nesrine F.

A mes collègues : Imane, Chaima, Wiam, Mouhamed Amine, Houssein,

Noufel, Saad, Nour Eddine, Diae eddine.

À mes oncles et tantes et leurs enfants et toute la famille Ben Mouhoub.

*Housna*

## Résumé

L'extraction d'informations s'intéresse à l'acquisition, l'organisation, le stockage et la recherche des informations, ainsi qu'au développement des techniques, des méthodes et outils permettant de retrouver les informations pertinentes qui assurent la satisfaction des besoins de l'utilisateur dans un langage naturel. Pour être plus précis, les systèmes d'extraction d'information doivent exploiter des ressources sémantiques, notamment des ontologies.

L'ontologie est une spécification explicite et formelle d'une conceptualisation qui joue un rôle majeur dans le processus d'extraction. L'extraction d'information à base d'ontologie (OBIE) est différente de l'EI traditionnelle du fait qu'elle identifie les types des entités extraites en les reliant à leur description sémantique dans l'ontologie.

Dans ce contexte, nous avons proposé une approche d'extraction d'information basée sur l'ontologie à partir des entités nommées, le but de ce travail est de présenter certaines techniques pour améliorer l'extraction d'information à partir des entités nommées. L'objectif est d'utiliser les entités nommées dans l'extraction de l'information est de renforcer la performance des résultats donnés et améliorer leur qualité.

Dans cette étude nous avons présenté notre approche d'un système d'extraction d'information sur le web dans le domaine de catastrophes naturelles afin d'extraire des entités nommées Spatiales et Caractéristiques de catastrophe naturelle et les dégâts. Nous avons cité les différents résultats avant et après l'ajout des règles JAPE et à base des ontologies.

**Mots clés :** Extraction d'Information, Entité Nommée, Ontologie, JAPE, Catastrophe Naturelle.

## **Abstract**

Information extraction is concerned with the acquisition, organization, storage and retrieval of information, as well as the development of techniques, methods and tools for retrieving relevant information that satisfies the user's needs in a natural language. To be more precise, information extraction systems must exploit semantic resources, especially ontologies.

The ontology is an explicit and formal specification of a conceptualization that plays a major role in the extraction process. Ontology-based information extraction (OBIE) is different from traditional IE in that it identifies the types of extracted entities by linking them to their semantic description in the ontology.

In this context, we have proposed an ontology-based information extraction approach from named entities, the goal of this work is to present some techniques to improve information extraction from named entities. The objective of using named entities in information extraction is to enhance the performance of the given results and improve their quality.

In this study we presented our approach of a web-based information extraction system in the field of natural disasters in order to extract Spatial Named Entities and Natural Disaster Characteristics and damages. We cited the different results before and after the addition of JAPE rules and based on ontologies.

**Keywords :** Information Extraction, Named Entity, Ontology, JAPE, Natural Disasters.

## ملخص

استخراج المعلومات يهتم بالحصول على المعلومات وتنظيمها وتخزينها واسترجاعها، فضلاً عن تطوير التقنيات والأساليب والأدوات لاسترجاع المعلومات ذات الصلة التي تضمن تلبية احتياجات المستخدم باللغة الطبيعية. لكي تكون أكثر دقة ، يجب أن تستغل أنظمة استخراج المعلومات الموارد الدلالية ، ولا سيما الأنطولوجيا.

الأنطولوجيا هو تحديد صريح ورسمي للتصور الذي يلعب دوراً رئيسياً في عملية الاستخراج. يختلف استخراج المعلومات القائم على علم الوجود (OBIE) عن IE التقليدي من حيث أنه يحدد أنواع الكيانات المستخرجة من خلال ربطها بالوصف الدلالي في علم الوجود.

في هذا السياق ، اقترحنا أسلوب استخراج المعلومات القائم على الأنطولوجيا من الكيانات المسماة ، والهدف من هذا العمل هو تقديم بعض التقنيات لتحسين استخراج المعلومات من الكيانات المسماة. الهدف من استخدامها في استخراج المعلومات هو تعزيز أداء النتائج المعينة وتحسين جودتها.

في هذه الدراسة قدمنا نهجنا الخاص بنظام استخراج المعلومات على شبكة الإنترنت في مجال الكوارث الطبيعية من أجل استخراج الكيانات المحددة المكانية وخصائص الكوارث الطبيعية والأضرار. لقد استشهدنا بالنتائج المختلفة قبل وبعد إضافة JAPE والقواعد المستندة إلى علم الوجود.

**الكلمات الرئيسية:** استخراج المعلومات، الكيان مسمى، الأنطولوجيا ، JAPE ، الكوارث الطبيعية

# Table des Matières

Remerciement.....	i
Dédicaces .....	ii
Dédicaces .....	iii
Résumé .....	iv
Table des matières .....	vii
Liste des figures .....	x
Liste des tableaux .....	xi
Liste des sigles et acronymes .....	xii
Introduction générale.....	1
<b>Chapitre 01: Extraction de l'information.....</b>	<b>3</b>
1.1 Introduction .....	4
1.2 Extraction de l'information .....	4
1.2.1 Définition .....	4
1.2.2 Information.....	5
1.3 Principe de l'Extraction d'Information .....	5
1.4 Approches d'extraction d'information .....	5
1.4.1 Approches basée sur des règles .....	6
1.4.2 Approches d'apprentissage automatique.....	6
1.4.3 Approches hybrides.....	7
1.4.4 Approches basées sur l'ontologie .....	7
1.5 La structure générale d'un système d'Extraction de l'Information.....	8
1.5.1 Prétraitements.....	8
1.5.2 Analyse linguistique.....	8
1.5.3 Installation des formulaires .....	8
1.6 Recherche et Extraction d'Information .....	9
1.6.1 Recherche d'Information (définition) .....	9
1.6.1.1 Modélisation des histoires et des requêtes .....	10
1.6.1.2 Couplage.....	10
1.6.1.3 Production et mise en forme des résultats.....	10
1.6.2 Différences et liens avec l'Extraction d'Information .....	11
1.6.3 Relation entre la recherche et l'extraction d'information .....	12
1.7 Conclusion.....	12
<b>Chapitre 02: Entités Nommées et les ontologies .....</b>	<b>14</b>
2.1 Introduction .....	15
2.2 Entités Nommées.....	15
2.2.1 Définition .....	15
2.2.2 Formes des entités nommées.....	15
2.2.3 Types des entités nommées .....	16
2.2.4 Le rôle d'entités nommées .....	17
2.2.5 Reconnaissance des entités nommées .....	17
2.2.6 Catégorisation.....	17
2.2.7 Extraction des entités nommées (EEN).....	18
2.2.7.1 Extraction des informations du texte.....	18



2.2.7.2 Répondre automatiquement à des questions .....	18
2.2.7.3 Annotation manuelle de corpus des entités nommées.....	19
2.3 L'extraction d'information basée sur les ontologies.....	20
2.3.1 Définition de l'ontologie .....	20
2.3.2 Les types d'ontologies .....	21
2.3.2.1 Ontologie de représentation de connaissances .....	21
2.3.2.2 Ontologie de haut niveau / supérieure (Top-level / Upper-model) .....	21
2.3.2.3 Ontologie générale (Généricontologie).....	22
2.3.2.4 Ontologie de domaine .....	22
2.3.2.5 Ontologie d'application (Application ontology).....	22
2.3.3 Domaines d'applications des ontologies .....	22
2.3.4 L'extraction d'information basée sur les ontologies .....	24
2.4 La relation entre les entités nommées et l'ontologie .....	27
2.5 Conclusion.....	28
<b>Chapitre 03: Conception du système d'extraction d'information à base d'ontologie .....</b>	<b>29</b>
3.1 Introduction .....	30
3.2 Travaux reliés.....	30
3.2.1 KIM .....	30
3.2.2 SPRAT .....	31
3.2.3 SPEED .....	31
3.3 Notre Approche proposée.....	32
3.4 Architecture de modèle proposé .....	33
3.5 Étapes de l'approche proposée .....	34
3.5.1 Étape Collection des données .....	34
3.5.2 Étape création de l'ontologie .....	34
3.5.3 Étape d'analyse d'un texte .....	35
3.5.3.1 Tokeniser.....	35
3.5.3.2 sentence splitter .....	36
3.5.3.3 Part Of Speech Tagger .....	36
3.5.3.4 Analyseur Morphologique.....	37
3.5.3.5 OntoRoot Gazetteer.....	37
3.5.3.6 SemanticTagger.....	38
3.5.4 Algorithme .....	39
3.6 Conclusion.....	40
<b>Chapitre 04: Implémentation .....</b>	<b>41</b>
4.1 Introduction .....	42
4.2 Protégé.....	42
4.2.1 Création de l'ontologie.....	43
4.3 GATE .....	46
4.3.1 ANNIE .....	47
4.3.2 Le formalisme JAPE .....	48
4.3.3 Ontrootgazeteer .....	48
4.4 Différentes Méthodes Appliquées .....	50
4.4.1 La première méthode Appliquée .....	50
4.4.2 La Deuxième Méthode Appliquée .....	52
4.4.3 La Troisième Méthode Appliquée.....	54

4.5 Evaluation.....	56
4.6 Conclusion .....	58
Conclusion générale .....	59
Bibliographie .....	60
Annexes.....	63

# Liste des figures

<b>Figure 1.1</b> : Schéma générale de la tâche d'extraction d'information .....	4
<b>Figure 1.2</b> : Le Architecture générale d'un système OBIE.....	7
<b>Figure1.3</b> : Architecture générale d'un système d'EI .....	9
<b>Figure 1.4</b> :Le processus de la recherche d'information .....	10
<b>Figure 1.5</b> :La recherche d'information et L'extraction d'information .....	11
<b>Figure2.1</b> :Modèle du système de réponse automatique fourni des notes de concours.....	19
<b>Figure 2.2</b> :Éléments d'un processus d'annotation .....	20
<b>Figure 2.3</b> : Exemple d'un petite ontologie .....	21
<b>Figure2.4</b> : Web sémantique.....	24
<b>Figure2.5</b> : Système OBIE .....	25
<b>Figure 2.6</b> : Extraction d'informations basée sur l'ontologie.....	25
<b>Figure 2.7</b> : Architecture générale du système OBIE.....	27
<b>Figure 3.1</b> : KIM Architecture présentée .....	30
<b>Figure 3.2</b> : Architecture SPEED telle que présentée .....	32
<b>Figure 3.3</b> : Architecture générale du système proposé .....	34
<b>Figure 3.4</b> : Building Ontology Resource Root (OntoRoot) Gazetteer from the Ontology .....	38
<b>Figure3.5</b> :Exemple de règle avec le formalisme JAPE.....	39
<b>Figure 4.1</b> : Protégé Screenshot .....	42
<b>Figure 4.2</b> : Classes d'ontologie de haut niveau .....	44
<b>Figure 4.3</b> : Hiérarchie de propriété d'objet.....	45
<b>Figure 4.4</b> : Liste de certaines instances d'ontologie .....	46
<b>Figure 4.5</b> : Exemple de l'interface dans GATE .....	47
<b>Figure 4.6</b> : Parametres de l'ontorootgazeteer .....	49
<b>Figure4.7</b> : Les PRs de l'application rootfinder .....	50
<b>Figure 4.8</b> : Exécution d'entités nommées spatial avant les règles JAPE .....	51
<b>Figure 4.9</b> : Exécution d'entités nommées (caractéristiques) avant les règles JAPE .....	51
<b>Figure 4.10</b> : Exécution d'entités nommées(Dégâts) avant les règles JAPE .....	52
<b>Figure 4.11</b> : Exécution d'entités nommées spatial après les règles JAPE.....	53
<b>Figure 4.12</b> : Exécution d'entités nommées (caractéristiques) après les règles JAPE .....	53
<b>Figure 4.13</b> : Exécution d'entités nommées(Dégâts)après les règles JAPE .....	54
<b>Figure 4.14</b> : Exécution d'ontologie (lookup )(location) .....	55
<b>Figure 4.15</b> : Exécution avec l'ontologie (lookup caractéristiques).....	55
<b>Figure 4.16</b> : Exécution a base d'ontologie (lookup).....	56

## Liste des tableaux

<b>Table 2.1 :</b> Exemple d'entités nommées .....	16
<b>Table 4.1:</b> Ontology Classes and Sub-classes .....	44
<b>Table 4.2:</b> Numéros d'entité nommés extraits de documents .....	57
<b>Table 4.a:</b> Evaluation : Entité Nommée Spatiale .....	57
<b>Table 4.b:</b> Evaluation : Entité Nommée Dégâts des catastrophes naturelles .....	57
<b>Table 4.c:</b> Evaluation : Entité Nommée Caractéristiques des catastrophes naturelles .....	57

## Liste des sigles et acronymes

- ANNIE** A Nearly New Information Extraction
- CRF** Conditional random fields
- EEN** Extraction des Entités Nommées
- EN** Entités Nommées
- EI** Extraction d'Information
- GATE** General Architectur for Text Engineering
- JAPE** Java Annotation Patterns Engine
- KIM** Knowledge and Information Management
- MHV** mots hors vocabulaire
- MUC** Message Under Standing Conférences
- NER** Named Entity Recognition
- NDO** Natural Disasters Ontology
- OBIE** Ontology Based Information Extraction
- OWL** Web ontology Language
- SPEED** Semantics-Based Pipeline for Economic Event Detection
- SPRAT** Semantic Pattern Recognition and Annotation Tool
- TAL** Traitement Automatisées des Langues
- TALN** traitement automatique des langages naturel

# Introduction

---

## Introduction générale

Les systèmes d'information sont de plus en plus accessibles via Internet ou Intranet. Ils permettent aux utilisateurs d'accéder à une grande quantité d'informations d'une ou plusieurs sources. Cette croissance rapide pose la question de savoir comment trouver les informations qui nous intéressent dans ce grand groupe des données.

En effet, les besoins d'information sont essentiels dans de nombreux domaines, comme la recherche ou la veille scientifique et technologique, mais avec la croissance exponentielle de l'information électronique sur le web, la satisfaction de besoins d'information des utilisateurs reste un objectif imperceptible et difficile.

L'extraction automatique de l'information est un processus important et nécessaire lorsque l'information est devenue un fichier électronique, sur les sites web, les médias ou les réseaux sociaux. Cependant l'extraction d'informations sans l'utilisation de moyens auxiliaires ou la spécification du contenu de ce que l'utilisateur souhaite précisément, ne donne pas de résultats satisfaisants et efficaces, puisqu'il est possible de se trouver dans certains cas de similitude ou d'erreur totale.

Pour résoudre ce problème, le système d'extraction basée sur l'ontologie (OBIE) est différent de l'IE traditionnelle, car il identifie le type d'entités extraites en les reliant à leurs descriptions sémantique dans les ontologies.

Dans cette proposition, le but de ce travail est améliorer la recherche d'information. Notre principale approche porte sur l'identification des entités nommées contenues dans un ensemble de documents ou de textes porté en ligne.

Pour développer cet aspect nous avons subdivisé notre travail en quatre chapitres :

- Dans le premier chapitre, nous introduisons des notions générales sur l'extraction de l'information ou nous allons présenter la définition générale de l'extraction d'information, le principe et ainsi que le processus, puis nous définissons les différentes méthodes d'extractions et l'architecture générale d'EI. Enfin, nous expliquons la relation entre l'information et l'extraction d'information.
- Dans le deuxième chapitre, nous parlons des entités nommées, des différentes formes, types d'entités nommées, catégorisation et rôle. Pour la partie ontologies, nous allons évoquer la définition des ontologies, différents types, domaines et les différents systèmes d'extraction basés sur l'ontologie. Nous terminons par l'explication d'une architecture générale d'un système d'extraction de l'information basé ontologie.

## **Introduction**

---

- Dans le troisième chapitre, nous présentons des travaux reliés, en proposant et expliquant une approche en détail. Ensuite, nous présentons l'architecture de l'approche proposée, ses étapes et quelques exemples illustratifs.
- Dans le quatrième chapitre, nous allons définir l'outil Protégé pour la création de l'ontologie, le GATE développer et ces composants ANNIE et le formalisme JAPE pour l'extraction. Nous expliquons par la suite trois méthodes et les différentes étapes utilisées dans notre application dans GATE ainsi que l'évaluation de chaque méthode.

# **Chapitre 1 :**

## Extraction de l'information



## 1.1 Introduction

Le développement et l'évolution rapide de la technologie et l'avancement des outils de production d'informations tels que les éditeurs de texte et les réseaux sociaux créent une quantité énorme d'informations chaque jour .

La croissance rapide du volume d'informations engendre des problèmes de trouver et extraire des informations pertinentes et qui nous intéressent. L'extraction d'information est une technologie récente et moderne mais elle répond à un besoin très ancien : acquérir la connaissance à partir du texte. Elle devient nécessaire beaucoup plus avec l'essor considérable de la masse de document électronique (courrier électronique, Internet, etc.). Dans ce chapitre, nous présenterons des notions de base sur l'extraction de l'information.

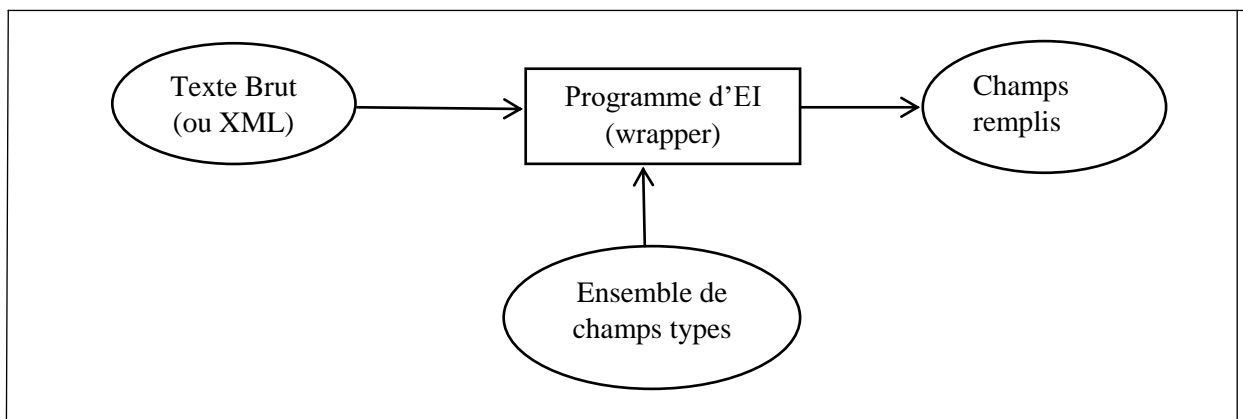
## 1.2 Extraction de l'information

### 1.2.1 Définition

L'extraction d'informations est une technologie moderne qui extrait et organise automatiquement les informations pertinentes d'un ou plusieurs documents en langage naturel. L'extraction s'effectue au moyen de formulaires prédéfinis, ces modèles sont définis pour les informations requises à travers une structure prédéfinie tels que : les entités, les relations entre elles et les événements impliqués dans ces entités [1].

Ce n'est qu'avec l'avènement de MUC-7 (Message Under standing Conférences) que le domaine de l'exploration de l'information a fait des progrès significatifs [2].

La figure suivante représente un schéma générale de la tâche d'extraction d'information



**Figure 1.1:** Schéma générale de la tâche d'extraction d'information [3].

**1.2.2 Information**

L'information est un concept aux multiples significations. Elle est étroitement liée à des concepts tels que la contrainte, la communication, le contrôle, les données, la forme, l'instruction, la connaissance, le sens, la perception et la représentation. Au sens étymologique, l'information est tout ce qui façonne l'esprit. Elle vient du verbe latin *informare*, qui signifie « former » ou « former une idée ». L'information spécifie à la fois l'information à transmettre et les symboles utilisés pour l'écriture en utilisant des codes avec des symboles significatifs tels que des alphabets, des bases numériques, des idéogrammes ou des pictogrammes. [4].

**1.3 Principe de l'Extraction de l'Information**

L'extraction implémente une analyse du texte pour interpréter et construire une représentation officielle qui fournira automatiquement des réponses précises à l'utilisateur. Il ne suffit pas de choisir un fragment brut de texte mais de mettre des éléments en relation avec la restauration d'une information complète et structurée sauf dans des cas très simples.

C'est une tâche compliquée qui exige de comprendre et de connaître, les ressources linguistiques, les sémantiques et les concepts adaptés à la documentation et au champ à traiter.

Les techniques de recherche documentaire fondées sur les technologies de traitement automatisées des langues (TAL) permettent de structurer les informations textuelles qui, au départ, manquent de structure logique. Le domaine de l'EI est souvent décomposé en plusieurs sous problèmes qui sont :

- L'extraction d'entités nommées ;
- L'extraction de descripteurs thématiques (libres ou normalisés) ;
- L'extraction de phrases importantes sous un point de vue donné ;
- L'extraction d'attributs ;
- L'extraction d'associations entre entités nommées et descripteurs ;
- L'extraction de correspondances multilingues [5].

**1.4 Approches d'extraction d'information**

Les systèmes d'extraction de l'information reposent généralement sur deux approches principales : l'approche dite à base de règle et l'approche à base d'apprentissage

automatique. Il existe des systèmes hybrides qui combinent ces deux approches et il y'a aussi des méthodes à base d'ontologie qui sont expliquées comme suit :

#### 1.4.1 Approches basée sur des règles

Elles reposent sur l'identification d'entités à l'aide de règles construites manuellement à partir d'un corpus d'un domaine donné. Les méthodes basées sur des règles, également appelées méthodes d'ingénierie des connaissances dans certaines sources, fonctionnent bien, mais elles nécessitent encore beaucoup d'efforts et de temps pour analyser les données et écrire des règles. Les systèmes basés sur la construction manuelle des règles sont plus intéressants dans des domaines fermés où l'intervention humaine est à la fois indispensable et accessible.

Dans des domaines ouverts telle que l'extraction d'opinions à partir de blogs, la flexibilité des méthodes statistiques est plus appropriée [2].

#### 1.4.2 Approches d'apprentissage automatique

Les méthodes d'apprentissage automatique ou bien les méthodes d'apprentissage statique, sont des techniques capables d'extraire obligatoirement, par encyclique, des grossiers étiquetés dans lequel un jeu de comptables.

Il existe différents types d'apprentissage : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement [6].

- **Apprentissage supervisé**

L'apprentissage supervisé est s'appuyé sur des données ou exemples labellisés (étiquetés ou annotés) pour entraîner des modèles d'intelligence artificielle (IA) prédictifs.

Il est appliqué lorsque les données sont dans le formulaire des variables d'entrée et des valeurs cibles de sortie [6].

- **Apprentissage non supervisé**

L'apprentissage non supervisé mise en œuvre lorsque les données sont disponibles uniquement sous la forme d'une entrée et il n'y a pas de variable de sortie correspondante.

Ces algorithmes modélisent le sous-jacent des caractéristiques des données afin d'en savoir plus sur leurs caractéristiques [6].

- **Apprentissage par renforcement**

Le « renforcement Learning » ou apprentissage par renforcement, est une méthode de machine learning permettant de réaliser des tâches complexes de façon autonome. Lorsqu'elles ont un résultat positif et induisent des récompenses, on conclut que ces expériences sont positives et qu'elles doivent être recommencées. Inversement, si le

résultat de l'expérience n'est pas concluant, on le mémorise pour ne plus faire la même erreur [7].

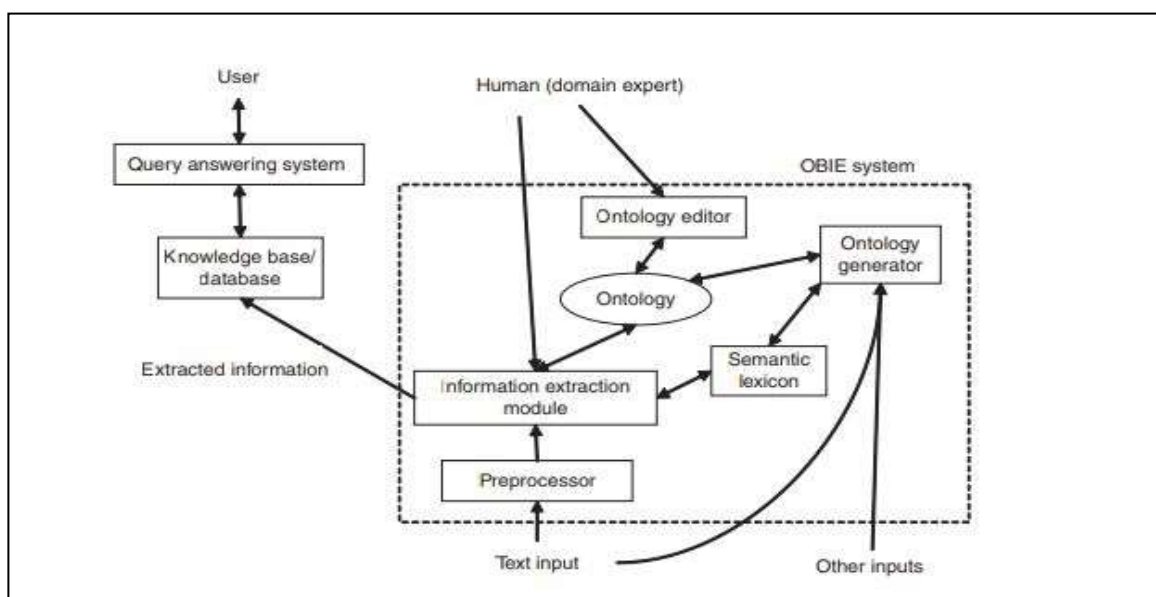
### 1.4.3 Approches hybrides

L'approche hybride utilise des règles manuscrites, mais certains de leurs règles utilisés construit également des informations syntaxiques et discursives extraites des données à l'aide d'algorithmes d'apprentissage automatique.

La combinaison des deux méthodes peut être effectuée par la méthode manuel basée sur des règles où l'hybridation est effectuée en convertissant les résultats de base des règles puis considérer ces résultats comme des attributs pour la classification CRF, ou par une méthode d'apprentissage où l'hybridation est faite par utilisation de l'algorithme d'apprentissage des règles pour la détection des entités nommés, par la suite les résultats obtenus sont validés par un groupe de règles construites manuellement en calculant un score de confirmation pour chaque règle [8].

### 1.4.4 Approches basées sur l'ontologie

L'extraction d'informations basée sur des ontologies (OBIE : Ontology Based Information Extraction) est apparue comme un sous-domaine de l'extraction d'informations. Les ontologies sont utilisées par le processus d'extraction d'informations et la sortie est généralement présentée à travers une ontologie. Il convient de noter qu'une ontologie se définit comme une spécification formelle et explicite de conceptualisations partagées.



**Figure 1.2:** L' Architecture générale d'un système OBIE [9]

En règle générale, une ontologie est spécifiée pour un domaine spécifique. Étant donné que l'extraction d'informations consiste essentiellement à récupérer des informations pour un domaine particulier, une spécification formelle et explicite des concepts du domaine par le biais d'une ontologie peut être utile pour le processus.

Par exemple, les ontologies géopolitiques qui définissent des concepts tels que le pays, la province et la ville peuvent être utilisées pour guider le système d'extraction d'informations décrit précédemment. C'est l'idée générale derrière l'extraction d'informations basée sur l'ontologie [10].

### **1.5 La structure générale d'un système d'Extraction de l'Information**

La structure générale de la figure 1.3 montre les composants importants de l'extraction d'informations, qui se déroule en trois phases [11] :

#### **1.5.1 Prétraitements**

Elle consiste en un ensemble d'opérations « de surface » sur un matériau linguistique qui permet l'analyse linguistique avec texte "nettoyé" et "préparé".

#### **1.5.2 Analyse linguistique**

Dans cette phase l'analyse morphologique consiste à étiqueter les mots conformément à eux (nom, verbe, adjectif,...) et à percevoir leur genre, nombre, personne, etc. ensuite l'analyse syntaxique doit produire (de manière plus ou moins exhaustive) les relations grammaticales : relations sujet-verbe ou verbe-COD, rattachements prépositionnels (compléments de nom ou compléments indirects des verbes), etc.

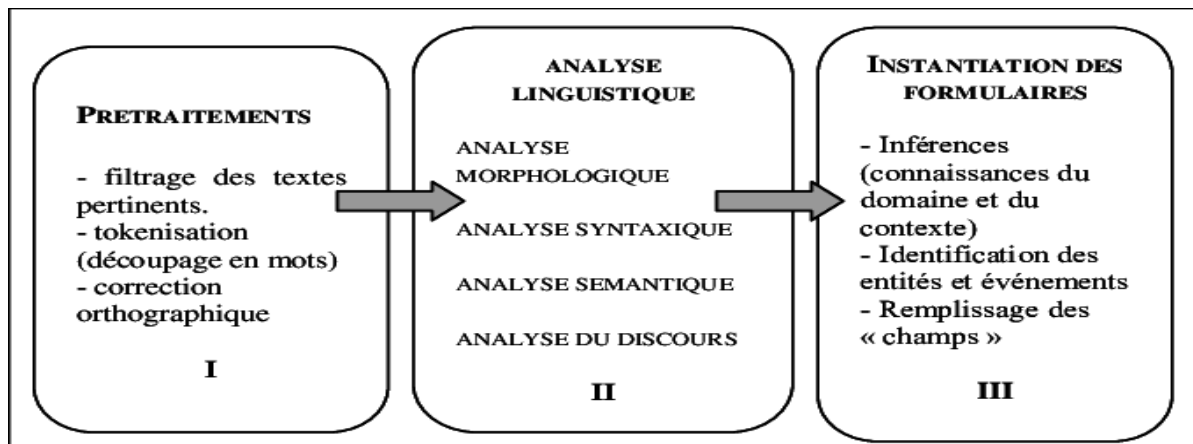
L'analyse sémantique vise à construire à partir de chaque proposition une « représentation conceptuelle » sous forme d'expression logique ou de réseau sémantique.

Enfin l'analyse du discours doit établir les liens entre les différentes phrases, spécifiquement repérer les coréférences ou l'ordre temporel des énoncés. L'organisation de ces traitements, et bien sûr les méthodes linguistiques utilisées constituent des caractéristiques importantes des différents systèmes d'EI [11].

#### **1.5.3 Installation des formulaires**

Dans cette phase il s'agit de remplir les champs des formulaires. Ce qui suppose particulièrement d'affiner le calcul de l'identification des entités et des événements. Du point de vue des méthodes, une caractéristique de cette étape est d'être complètement orientée par le but, c'est à dire par la structure des formulaires, et de recourir à des

inférences mettant en jeu des connaissances du domaine (« le monde financier», « le monde de la route »...) alors que la phase d'analyse linguistique et elle plutôt orientée par la structure linguistique du texte [11].



**Figure1.3** : Architecture générale d'un système d'EI [11].

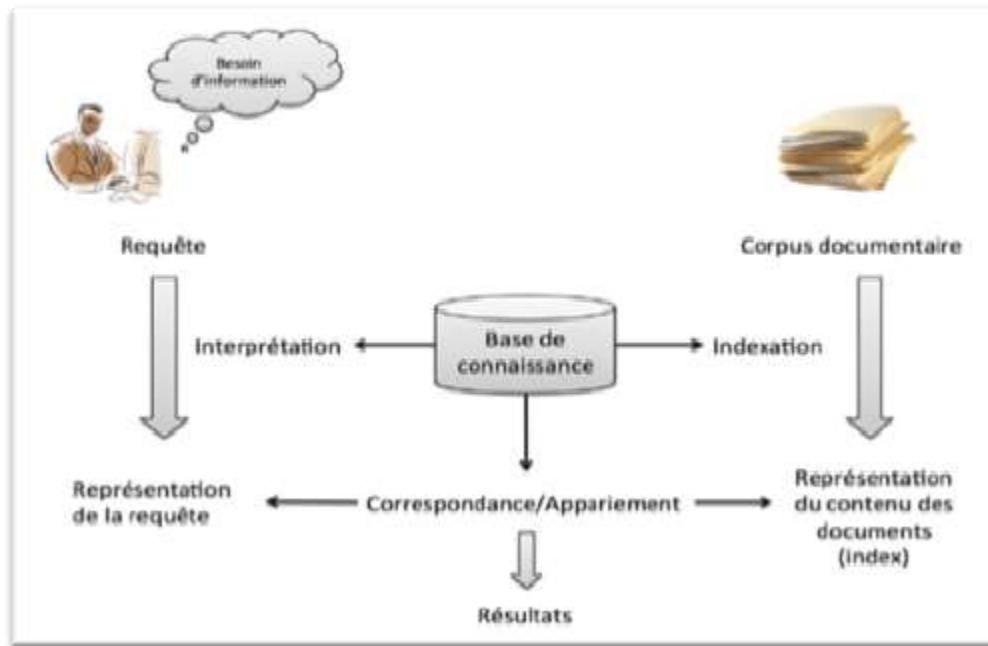
## 1.6 Recherche et Extraction d'Information

### 1.6.1 Recherche d'Information (définition)

La recherche d'informations est la science qui consiste à répondre de manière appropriée aux requêtes en trouvant des informations dans des groupes. Il s'agit de documents provenant d'une ou de plusieurs bases de données, décrits par un contenu ou des métadonnées connexes [8].

Les bases de données peuvent être relationnelles ou non structurées, telles que les bases de données connectées à des réseaux via des liens hypertextes dans le World Wide Web, Internet et les intranets. Le contenu du document peut être du texte, du son, des images ou des données. La recherche d'informations est un domaine historiquement lié aux sciences de l'information et à la bibliothéconomie qui vise à représenter des documents en créant des index à partir desquels récupérer des informations. L'informatique a développé des outils de traitement de l'information et de construction de représentations documentaires au fur et à mesure de leur indexation et de recherche d'information. La recherche d'information est un domaine interdisciplinaire, intéressant même les sciences cognitives [8].

Classiquement, le processus de recherche d'informations se déroule en trois étapes comme le montre la figure 1.4 :



**Figure 1.4:** Le processus de la recherche d'information

### 1.6.1.1 Modélisation des histoires et des requêtes

D'un côté, les documents de la collection de textes sont modélisés, et d'autre part, la requête de l'utilisateur est transformée en modèle selon la représentation choisie pour le document [8].

### 1.6.1.2 Couplage

La modélisation de requête est mise en correspondance avec la modélisation documentaire. Cette étape vise à déterminer la pertinence d'un document pour la requête afin de choisir les documents qui conviennent le mieux à la requête [8].

### 1.6.1.3 Production et mise en forme des résultats

En fonction de la tâche à effectuer :

- renvoi de tous les documents ou d'une sélection de documents dans l'ordre décroissant de leur pertinence ;
- renvoi des documents de manière simple ou accompagnés d'un indice de pertinence ;
- mise en évidence de l'information via, par exemple, la mise en valeur de certains termes (coloration, soulignement, etc.) [1].

1.6.2 Différences et liens avec l'Extraction d'Information

L'extraction d'information et la recherche d'information poursuivent le même objectif (trouver de l'information dans un ensemble de textes), mais leurs réponses et leurs moyens de mise en œuvre diffèrent.

Leur différence fondamentale est la nature des informations qu'ils renvoient. Un modèle de recherche d'informations qui est indépendant des informations à rechercher, le texte d'une collection de documents, puis sélectionne ceux qui traitent un sujet donné (le sujet représenté par la requête) et les fournit à l'utilisateur.

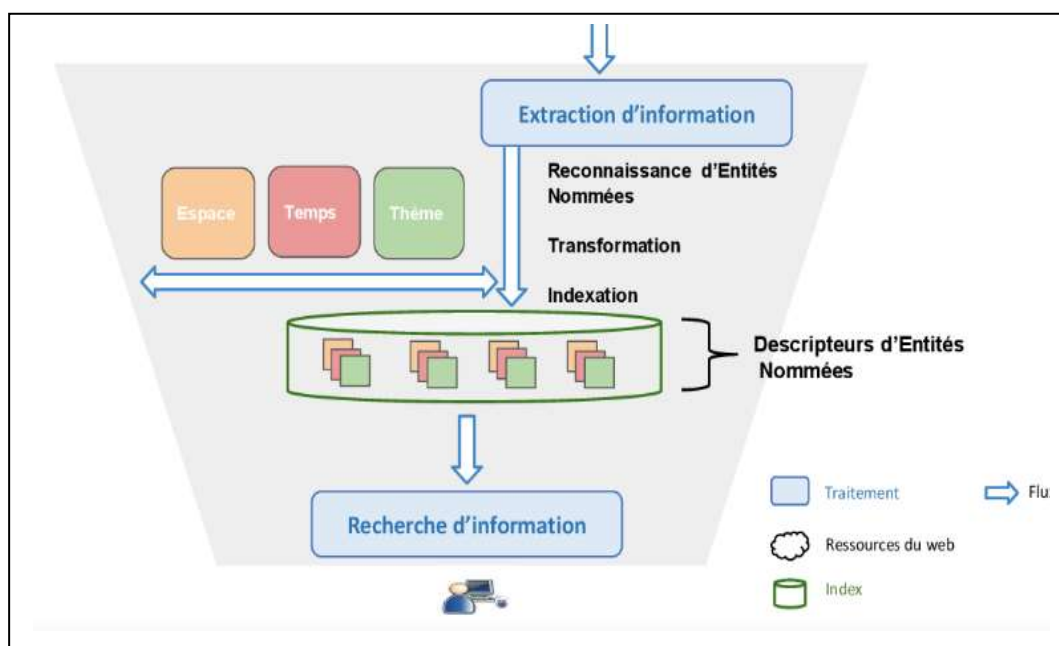


Figure 1.5: La recherche d'information et L'extraction d'information. [1]

Un tel système est ouvert (les demandes ne sont pas fixées a priori). Le système d'extraction d'informations analyse le document original pour n'en extraire que les informations précises qui intéressent l'utilisateur et qui sont précisées a priori (le système n'a pas de demande de saisie) [1].

Les deux méthodes utilisent des techniques différentes. Le travail des systèmes de recherche d'informations est influencé par la théorie de l'information, la théorie des probabilités et la théorie statistique, tandis que la recherche d'informations est dérivée de la recherche en linguistique informatique et en TALN (Traitement Automatique du Langage Naturel).



Les systèmes de recherche d'informations traitent généralement le texte comme une collection de mots non structurée. Au contraire, les systèmes d'extraction d'informations doivent s'intéresser à la structure grammaticale et aux propriétés syntaxiques du texte pour éviter des erreurs de sens importantes [1].

L'utilisation de système d'extraction d'information au lieu de la recherche d'information pour collecter des informations à partir d'un texte présente des avantages mais aussi des inconvénients : d'un côté, ils sont plus difficiles à mettre en œuvre et sont souvent liés à un domaine de connaissances spécifique, ce qui les rend difficilement adaptables à d'autres domaines, d'un autre côté, les résultats ne sont pas aussi précis que ce que donnerait un lecteur humain. Mais dans le cas de grands corpus, l'extraction d'information semble plus efficace que la recherche d'information, car la complexité et le coût de la tâche constituent la lecture et l'analyse manuelles d'un grand nombre de documents renvoyés par les systèmes de recherche d'informations, ces systèmes ne se sont pas couramment révélés suffisamment discriminants [1].

### **1.6.3 Relation entre la recherche et l'extraction d'information**

Il existe de nombreux moyens de fusionner ces deux systèmes :

- **Exploiter la Recherche d'Information en prétraitement de l'Extraction d'Information**

Face à un grand volume de textes, elle peut fournir à un système d'Extraction d'Information une sous-collection ne regroupant que les documents les plus pertinents.

- **Utiliser l'Extraction d'Information pour affiner les résultats d'un système de Recherche d'Information**

C'est l'amélioration de la phase de modélisation documentaire : les informations extraites de chaque document via un formulaire par un processus d'Extraction d'Information peuvent être utilisées pour créer un index qui modélise le document.

- **Prendre des techniques propres à l'Extraction d'Information**

Elle peut également servir à compléter les approches traditionnelles de recherche d'information pour catégoriser, filtrer, et classer les documents en fonction de leur pertinence.

## **1.7 Conclusion**

Dans ce chapitre nous avons présenté les concepts de base de l'extraction d'information. Nous avons mentionné la définition de l'extraction d'information, le

principe d'extraction d'information, les méthodes et l'architecture d'un système d'extraction d'information.

Nous avons également donné une comparaison entre la recherche d'information et l'extraction d'information.

Le chapitre suivant parle sur les entités nommées et les ontologies (définition, les types, l'utilisation) et l'extraction d'information basée sur les ontologies (Ontology Based Information Extraction).

# **Chapitre 2 :**

## Entités Nommées et les Ontologies

## 2.1 Introduction

La reconnaissance de l'entité nommée (NER) implique le traitement du texte et l'identification de certaines occurrences de mots appartenant à des catégories particulières des entités nommées. Elles sont depuis longtemps considérées comme un point central dans de multiples applications mettant en jeu des notions comme la compréhension et la recherche sémantique. L'extraction d'information doit utiliser des indicateurs pour préciser l'extraction d'information comme les entités nommées.

L'utilisation des ontologies dans la reconnaissance des entités nommée semble être un bon choix puisque nous pouvons sélectionner des ontologies spécifiques .

Dans ce chapitre, nous présenterons un ensemble de concepts de base nécessaires pour comprendre les entités nommées et les ontologies et extraire des informations basées sur eux.

## 2.2 Entités Nommées

### 2.2.1 Définition

Le concept d'Entité Nommée est apparu dans les années 90 au cours de la sixième conférence MUC (Message Understanding Conference), bien que l'EN (Entité Nommée) n'ait pas de définition standard, certains chercheurs ont proposé différentes définitions pour ce concept:

- «Les ENs sont des types d'unités lexicales particulières qui font référence à une entité du monde concret dans certains domaines spécifiques notamment humains, sociaux, politiques, économiques ou géographiques et qui ont un nom (typiquement un nom propre ou un acronyme)»
- «L'EN est un mot ou un groupe de mots désignant une personne, une organisation ou entreprise, un lieu, une date ou encore une expression numérique.» [8].

### 2.2.2 Formes des entités nommées

Il y a deux formes d'EN : les ENs simples et les ENs composées, chaque forme a un traitement différent :

- **Entités nommées simples**

Est une EN qui est constitué d'un seul mot, chez les noms de lieu « Jijel » et «Algérie » ou le nom de personne «Ali».

- **Entités nommées composées**

Est une EN qui est composée de deux ou plusieurs mots, comme par exemple le nom de personne « Abdel Hafid Boussof » et le nom de lieu «Ain Témouchent»

### 2.2.3 Types des entités nommées

Les entités nommées y sont réparties en 7 types primaires et 32 sous-types, dont voici la liste de sept types :

- **Personnes** : masses individuelles, individus collectives.
- **Lieux** : lieux administratifs (panorama, terroir, nations, surannations), lieux physiques (géographiques, hydrologiques, astrologiques).
- **Organisations** : industries, administrations.
- **Temps** : dates (absolues ou relatives) et horaires (absolus ou relatifs).
- **Montants** : exubérances, durées.
- **Produits** : objets travaillés, études artistiques, actions médiatiques, produits financiers, logiciels, récompenses, voies, doctrines, lois.
- **Impératifs** : offices individuelles, offices collectives.

Cette typologie ajoute donc aux entités nommées traditionnelles les produits et les fonctions. Elle ajoute une granularité supplémentaire (sous-types) aux types principaux (ou primaires)[1].

**Table 2.1** : Exemples d'entités nommées

Entité nommée	Type
Barack Obama	Nom de personne
USA	Nom de lieu
29 Décembre 2015	Expression temporelle
UNICEF	Nom d'organisation
18%	Expression pourcentage

### 2.2.4 Rôle d'entités nommées

Les ENs sont utiles pour développer des systèmes permettant de répondre à des questions, des résumés automatiques, la recherche d'informations, la traduction automatique (TA), le Web sémantique et la bio-informatique. Les EN sont également utilisés pour réduire l'incidence de l'utilisation abusive du vocabulaire (MHV) [10].

### 2.2.5 Reconnaissance des entités nommées

La plupart des systèmes de REN utilisent soit des approches de orientées connaissances soit des approches orientées données. Les systèmes orientées connaissances sont basée sur des lexiques (listes de prénom, de pays, etc.) et sur un ensemble de règles de réécriture. D'un part côté, les systèmes orientée données sont basés sur un modèle appris à partir d'un corpus préalablement annoté. Afin de profiter des avantages de ces deux approches, d'autres systèmes combinent des techniques d'apprentissage automatique et des règles produites manuellement [1].

La reconnaissance des entités nommées consiste à :

- Identifier des unités lexicales dans un texte ;
- Les catégoriser ;
- Eventuellement, les normaliser [13].

### 2.2.6 Catégorisation

Avec la richesse des informations qu'elles contiennent, les entités nommées (EN) sont des éléments très importants des systèmes de recherche d'informations. Dans les années 1980, l'activité d'évaluation MUC a défini la tâche de reconnaissance d'entité nommée.

Pour MUC-6, EN sont des noms propres, des acronymes et d'autres mots qui peuvent appartenir aux catégories suivantes :

- **Organisation** : comprend les entreprises, les agences gouvernementales et d'autres organisations ;
- **Personne** : regroupe les noms de personnes ou de familles ;
- **Localisation** : Regroupe les noms de lieux définis politiquement ou géographiquement (ville, pays, région, etc.) ;
- **Heure** : collecte les données de date et d'heure.

- **Numérique** : regroupe les données numériques, telles que les montants ou les pourcentages [12].

## **2.7 Extraction des entités nommées (EEN)**

L'Extraction des entités nommées (EEN) est la combinaison des méthodes pour indiquer des entités nommées dans des documents et les utiliser dans des objets différents. L'EEN a plusieurs applications dans la réalité. Dans cette partie, quelques applications dans des systèmes de traitement des informations automatiques par l'ordinateur sont présentées.[13]

### **2.7.1 Extraction des informations du texte**

Lors de l'extraction d'informations à partir de textes libres, à titre d'exemples d'informations d'une personne (nom, adresse, numéro de téléphone et emplacement de résidence), l'utilisateur doit lire les documents et noter toutes ces informations dans un tableau. Mais le travail est particulièrement avec le Big Data. Le système de récupération d'entité nommée peut retirer automatiquement ces informations. La phase d'extraction d'entités nommées consiste à mettre en place un système de détection et de typage des entités d'intérêt dans un texte [13].

### **2.7.2 Répondre automatiquement à des questions**

EEN (extraction d'entité nommée) joue un rôle important dans le système de réponse automatique. Le système peut connaître le nom de la personne et fournir les services correspondants.

La Figure 2.1 présente un exemple de système automatique fourni des notes de concours universitaire. Pour connaître les signes du concours, les gens doivent mettre en place un message en coordination constante (le message contient le numéro d'identification). Le système obtient ce numéro d'identification et trouve des notes dans la base de données. Mais si l'utilisateur ne connaît pas la forme par exemple, il n'a que le nom, la date de naissance ou d'autres informations, il est alors impossible de faire la recherche [13].

Dans ce système, l'utilisateur peut envoyer un message de n'importe quelle manière. Il contient les informations nécessaires (par exemple, nom, date de naissance, etc.). Le système reconnaît les entités nommées et les utilise pour trouver les notes correspondantes. Cependant,

EEN fait partie du système. Le système peut intégrer la reconnaissance vocale ou autre pour améliorer les fonctionnalités [8].

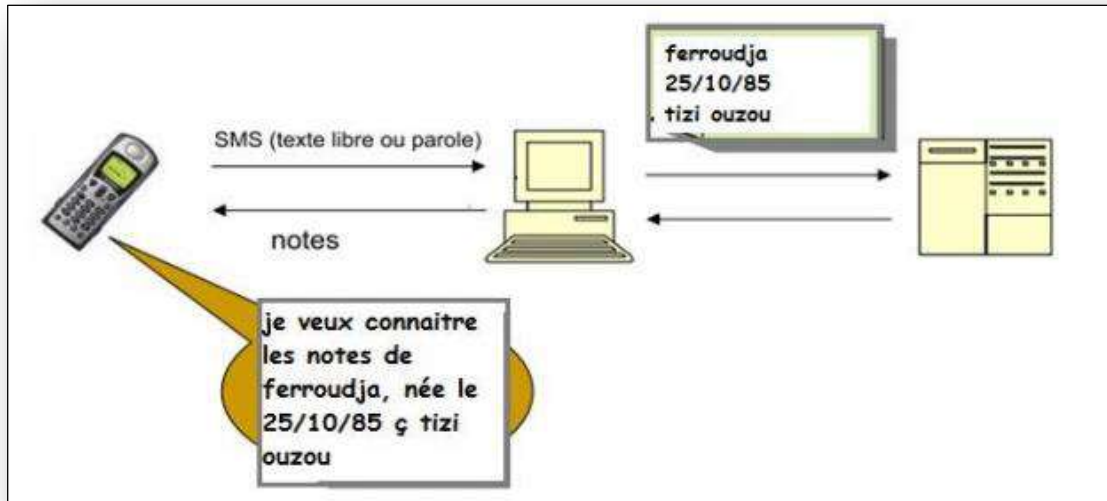


Figure 2.1 : Modèle du système de réponse automatique fourni des notes de concours [14].

### 2.7.3 Annotation manuelle de corpus des entités nommées

L'annotation de corpus (Figure 2.2) est une thématique très active qui fait l'objet de nombreux travaux. Effectivement, celle-ci peut être plus ou moins assistée, guidée et automatisée. De plus, le travail nécessite une grande rigueur et beaucoup de préparation afin d'obtenir une annotation fiable. Dans l'essentiel, trois éléments paraissent indispensables :

- **Guide d'annotation**

Il détaille les expressions linguistiques à annoter, selon des critères qui doivent laisser aussi peu de latitude que possible à la personne qui réalisera l'annotation.

- **Outils d'annotation**

Logiciels servant à annoter, dont les interfaces doivent faciliter, mais sans biaiser, le travail de l'annotateur, en incluant éventuellement une phase de pré-annotation automatique [10].



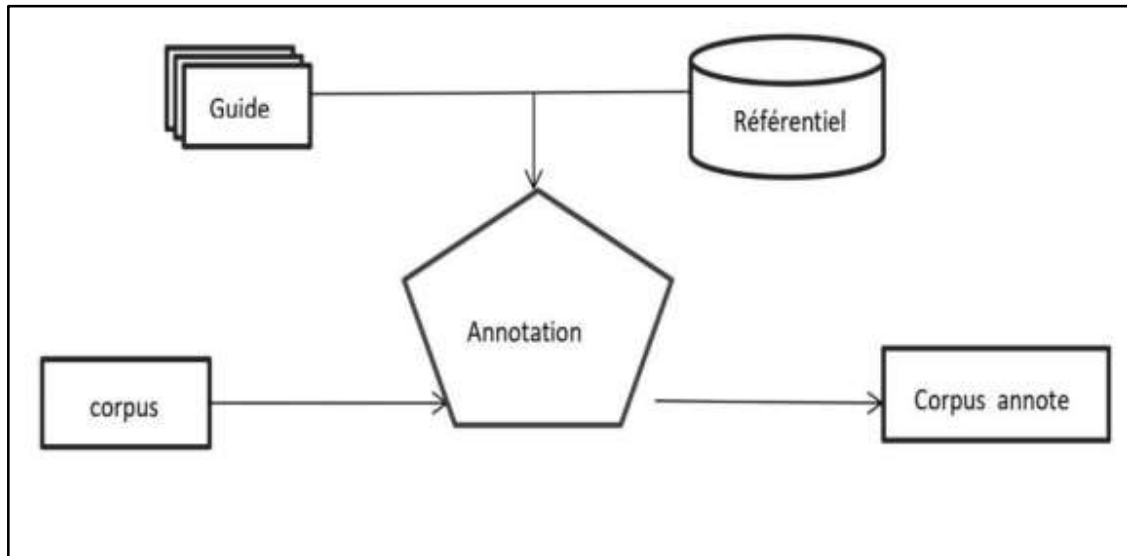


Figure 2.2 : Éléments d'un processus d'annotation [10]

## 2.3 Extraction d'information basée sur les ontologies

### 2.3.1 Définition de l'ontologie

Spécification explicite des conceptualisations partagées [15]. « formel » fait référence que l'ontologie doit être des faits compréhensibles par la machine. « Explicite » signifie que les types de concepts utilisés et les contraintes qui pèsent sur leur utilisation sont clairement définis. "Partagé" reflète l'idée de l'ontologie de capturer les connaissances consensuelles, c'est-à-dire qu'elles ne se limitent pas aux individus, mais sont acceptées par les groupes. La « conceptualisation » fait référence à un modèle abstrait d'un phénomène dans le monde, qui identifie les concepts connexes de ce phénomène [15].

Les ontologies peuvent être considérées comme des modèles déclaratifs qui définissent des domaines et représentent des concepts existants dans le domaine, leurs propriétés et les relations entre eux [14].

Elle joue un rôle essentiel dans le Web sémantique et tente de capturer la sémantique du domaine en déployant les connaissances brutes représentées, permettant aux machines de comprendre les relations entre les concepts d'un domaine. Il est généralement représenté sous la forme d'une base de connaissances qui devient alors disponible pour les applications qui ont besoin d'utiliser et/ou de partager la connaissance d'un domaine [15].

Les ontologies spécifient un ensemble des contraintes qui déclarent ce qui doit nécessairement tenir dans n'importe quel monde possible. Il avait l'habitude d'identifier ce qui « est » ou « pourrait être » dans le monde. L'objectif est de construire un modèle global complet pour décrire la sémantique de l'échange d'informations. En particulier dans le domaine de l'intelligence artificielle, les ontologies sont utilisées pour faciliter le partage et la réutilisation des connaissances. Les ontologies sont constituées de concepts, d'attributs, de relations et de contraintes entre les concepts. La figure 2.3 représente une ontologie simple, également appelée ontologie légère, contenant des classes et leurs relations taxonomiques.

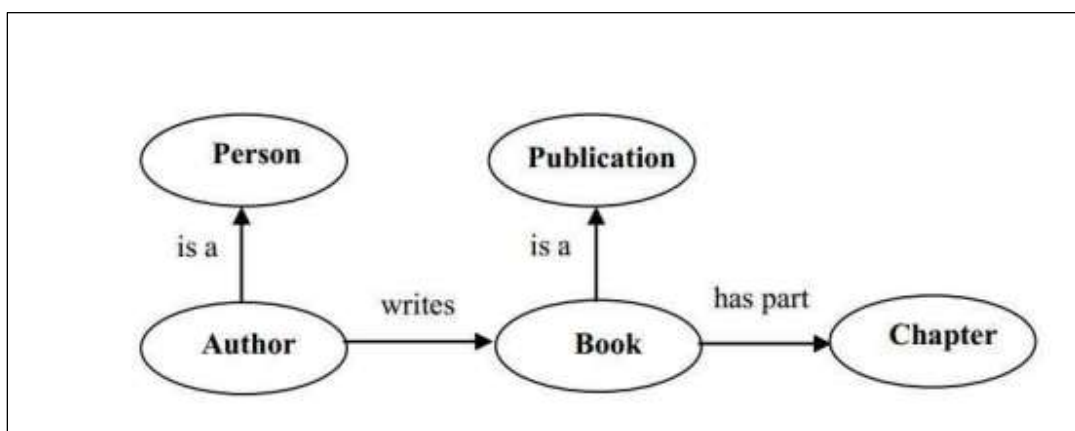


Figure 2.3: Exemple d'une petite ontologie [16]

### 2.3.2 Type d'ontologies

On distingue six types d'ontologie

#### 2.3.2.1 Ontologie de représentation de connaissances

Modéliser la représentation brute utilisée pour la formalisation des connaissances dans un paradigme donné. Par exemple, une ontologie sur le formalisme Topic Maps comprendrait les concepts suivants : Topic, Type of Topic, Association, Occurrence, Type Occurrence, etc. [18].

#### 2.3.2.2 Ontologie de haut niveau / supérieure (Top-level / Upper-model)

Elle exprime des concepts valables dans différents domaines. Elle a décrit certains concepts très généraux tels que l'espace, le temps, la matière, l'objet et l'événement, qui ne sont pas spécifiques à un problème ou à un domaine et doivent faire l'objet d'un large consensus communautaire, du moins dans la théorie des utilisateurs. Ce type d'ontologie est basé sur la théorie de la dépendance. Ses thèmes et l'étude des catégories de choses qui existent dans le

monde. En tant que concept très abstrait tel qu'entité, événement, état, processus, action, temps, espace, relation, propriété [18].

### **2.3.2.3 Ontologie générale (Generic ontologie)**

Également connue sous le nom d'ontologie centrale, la connaissance du modèle n'est pas aussi abstraite que celle véhiculée par les ontologies de haut niveau, mais elle est encore suffisamment générique pour être réutilisée dans différents domaines. Cette ontologie inclut un vocabulaire relatif aux choses, événements, temps, espace, causalité, comportement, fonction, etc.

### **2.3.2.4 Ontologie de domaine**

Cette ontologie exprime la conceptualisation d'un domaine spécifique, qui a plusieurs applications de ce domaine. Il fournit de

4s concepts et des relations couvrant le vocabulaire, les activités et les théories dans ces domaines. Selon Mzoguchi, une ontologie de domaine représente la connaissance du domaine ou de la tâche en cours d'exécution. Par exemple, dans le contexte de l'apprentissage en ligne, un domaine peut être la formation [18].

### **2.3.2.5 Ontologie d'application (Application ontology)**

C'est l'ontologie la plus concrète, elle contient des concepts qui dépendent d'un domaine et d'une tâche spécifiques, elle est concrète et non réutilisable. Ces concepts correspondent généralement aux rôles que jouent les entités du domaine lors de l'exécution d'activités spécifiques. Le problème ici est donc de mettre en relation des concepts associés à une tâche particulière afin de décrire son exécution [18].

### **2.3.2.6 Ontologie de Taches (Task ontology)**

L'ontologie de taches fournit un vocabulaire systématisé des termes employés pour résoudre des problèmes liés aux taches qui peuvent être ou non du même domaine. Elle fournit un ensemble de termes au moyen desquelles nous pouvons décrire généralement comment résoudre un type de problèmes. Elle inclut des noms, des verbes et des adjectifs génériques dans les descriptions de taches.[18]

## **2.3.3 Domaines d'applications des ontologies**

### **➤ Système d'information**

L'intégration des ontologies dans les systèmes d'information vise à réduire, voire éliminer les confusions conceptuelles et terminologiques aux points clés du système, et tend à

partager la compréhension pour améliorer la communication, le partage, l'interopérabilité et la réutilisation éventuelle, de sorte que des déclarations formelles puissent être utilisées pour caractériser une certaine connaissance des informations que le système gère et s'appuie sur la formalisation de ces représentations et de leurs significations pour automatiser les tâches de traitement de l'information [18].

L'ontologie se retrouve dans une large famille de systèmes d'information. Elle est utilisée pour :

- Décrire et traiter des ressources multimédia ;
- Assurer l'interopérabilité d'applications en réseaux ;
- Piloter des traitements automatiques de la langue naturelle ;
- Construire des solutions multilingues et interculturelles ;
- Permettre l'intégration des ressources hétérogènes d'information ;
- Vérifier la cohérence de modèles ;
- Permettre les raisonnements temporel et spatial ;
- Faire des approximations logiques ; etc.

Ces utilisations des ontologies se retrouvent dans de nombreux domaines d'applications tel que :

- Intégration d'information géographique ;
- Gestion de ressource humaine ;
- Aide à l'analyse en biologie, suivi médicale informatisé ;
- Commerce électronique ;
- Enseignement assisté par ordinateur ;
- Bibliothèque numériques.

➤ **Web sémantique**

Un courant particulièrement prometteur pour l'expansion des systèmes à base d'ontologies est celui du Web sémantique (**Figure2.4**). Il s'agit d'une extension du Web actuel, dans laquelle l'information se voit associée à un sens bien défini, améliorant la capacité des logiciels à traiter l'information disponible sur le Web.

L'annotation des ressources d'information du Web repose sur des ontologies, elles sont aussi disponibles et échangées sur le Web. Grâce au Web sémantique, l'ontologie a trouvé un

jeu de formalismes standards à l'échelle mondiale, et s'intègre dans de plus en plus d'applications Web, sans même que les utilisateurs ne le sachent. Cela se fait au profit des logiciels qui à travers les ontologies et les descriptions qu'elles permettent, peuvent proposer de nouvelles fonctionnalités exploitant les effets d'échelles du Web pour en améliorer les effets [18].

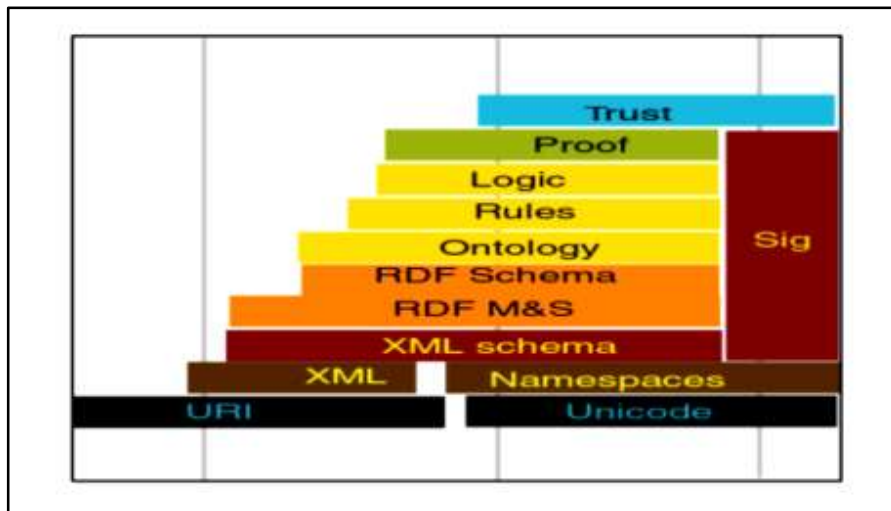


Figure2.4: Web sémantique [18]

### 2.3.4 Extraction d'information basée sur les ontologies (Ontology-based Information Extraction)

Le terme d'extraction d'information fondée sur l'ontologie (OBIE) (Figure 2.5) a récemment émergé en tant que sous-domaine d'extraction d'information. OBIE se distingue de l'EI traditionnelle par le fait qu'elle trouve le type d'entité extraite en l'associant à la description sémantique dans l'ontologie formelle. De plus, le processus d'extraction d'information utilise une ontologie et le résultat est habituellement présenté par l'ontologie.

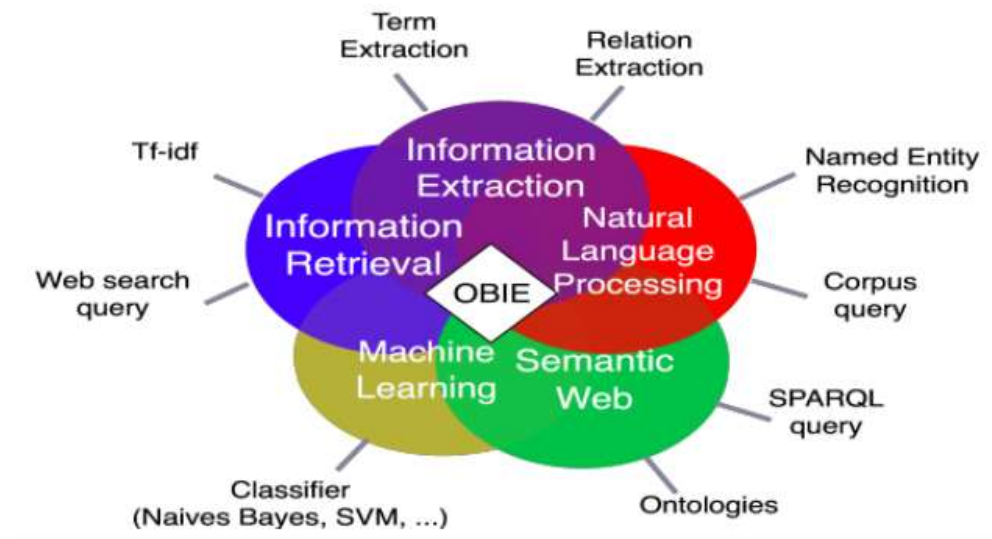


Figure2.5: système OBIE [20]

La définition précédente décrivait OBIE comme un système de traitement de texte en langage naturel non structuré ou semi-structuré qui utilise un mécanisme basé sur une ontologie pour extraire certains types d'informations et présenter la sortie à l'aide de l'ontologie. Ces systèmes peuvent être identifiés à un niveau supérieur. Dans de nombreux cas, les systèmes OBIE n'incluent pas ou n'utilisent pas tous ces composants [20].

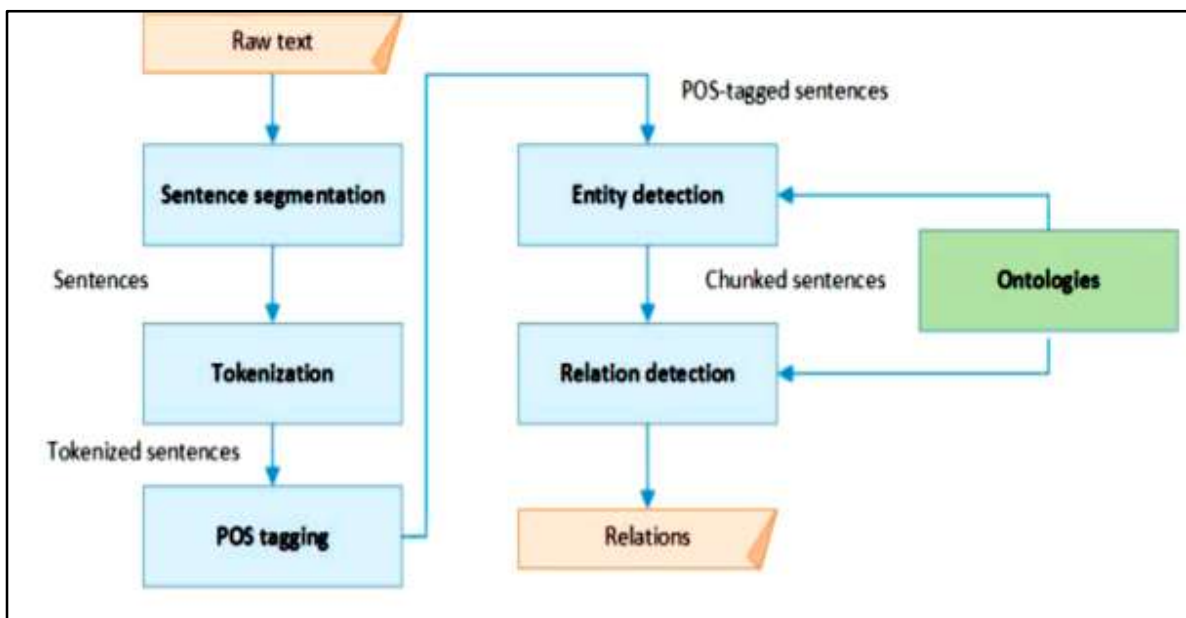


Figure 2.6 : Extraction d'informations basée sur l'ontologie.[20]

Au cours de l'introduction, les composants les plus courants sont identifiés, avec une attention particulière au traitement des données d'entrée textuelles. L'architecture est divisée en trois parties principales : les entrées, les modules et les sorties. A l'entrée, se trouvent les modules suivants :

- Les connaissances des experts du domaine, c'est-à-dire les connaissances des experts d'un domaine particulier, peuvent contenir différents types de sources de données : données structurées et semi-structurées, données non structurées souvent considérées comme l'entrée principale, et utilisateurs ;

L'architecture (**Figure 2.7**) contient également les modules suivants :

- Éditeurs de connaissances, générateurs de connaissances, préprocesseurs de texte, extracteurs d'informations, moteurs de recherche et moteurs de requête. Extracteurs d'informations, moteurs de recherche et réponders aux requêtes. Le module d'éditeur de connaissances comprend des outils de base pour maintenir la base de connaissances (par exemple, éditeur d'ontologie, moteur de recherche de dictionnaire, etc., tandis que le module de génération de connaissances fournit le référentiel de la base de connaissances).

L'objectif du module d'extraction d'informations est de détecter des instances de concept, des mots-clés, etc. à l'aide de textes annotés et de règles d'extraction.

Le module d'extraction d'information a pour but d'utiliser du texte annoté et des règles d'extraction pour détecter les instances conceptuelles, les attributs et les relations entre eux. quelle que soit la technique d'extraction d'information utilisée, elle est guidée par l'ontologie.

Les résultats de l'analyse des données non structurées sont stockés dans le module d'information d'extraction. Enfin, il convient de présenter les résultats de l'extraction textuelle de l'information sous une forme structurée. L'objectif d'une base de connaissances est de contenir des informations sous la forme d'un référentiel, véhiculant un type d'information [21].

En bref, les résultats du système OBIE incluent des informations extraites du texte. De plus, la sortie peut inclure des liens vers des documents texte à partir des quels les

informations ont été extraites. Document texte permettant d'extraire de l'information.

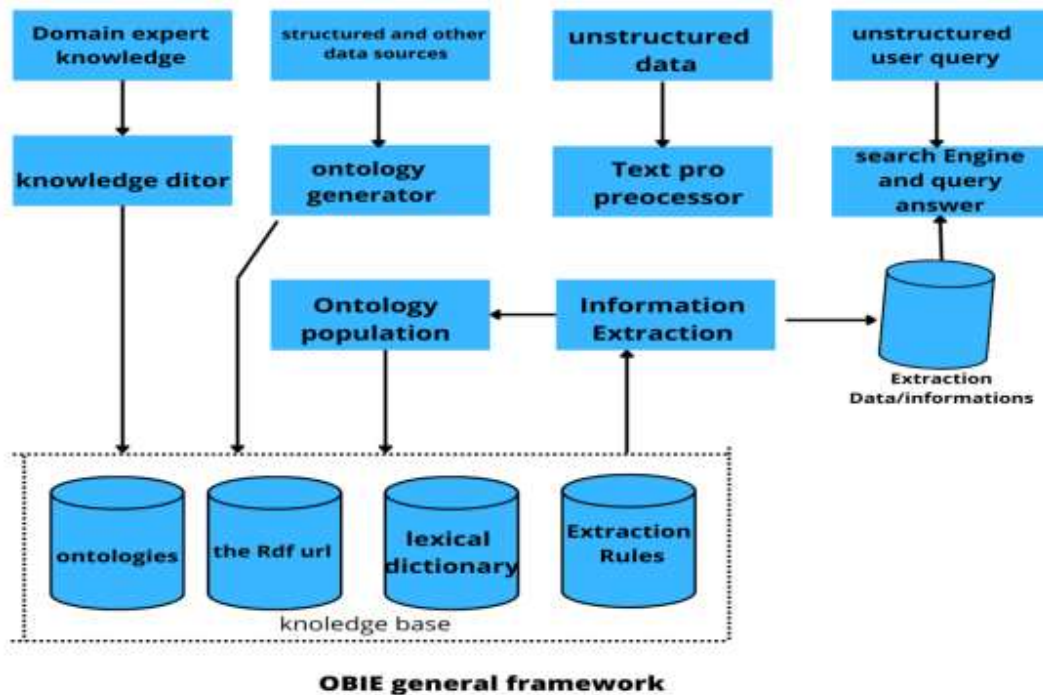


Figure 2.7 : Architecture générale du système OBIE [23]

Un système OBIE peut faire partie d'un système d'intervention plus large, où la sortie du processus OBIE est habituellement stockée dans une base de données ou de connaissances. Les moteurs de recherche et les mécanismes de réponse aux demandes explorent et utilisent les renseignements extraits, puis répondent aux questions des utilisateurs.

#### 2.4 Relation entre les entités nommées et l'ontologie

Ce référentiel vise à collecter les informations assemblées à partir des messages pour organiser et maintenir dans une structure cohérente afin de rendre son fonctionnement possible et facile de classer ou de chercher. Le modèle de représentation de ces connaissances est l'ontologie, elle conçue spécifiquement pour modéliser des classes conceptuelles correspondant à la classification classique des entités nommées (personnes, lieux, temps et événements). Le langage ontologique de choix est OWL. En plus des catégories de fonctionnalités, il y a d'autres concepts importants de modélisation dans cette ontologie.

Cependant, l'information est plus complexe et avec le temps peut conduire à une maintenance difficile d'une telle ontologie qui concerne dans les relations particulières entre



entités telles que l'appartenance d'une personne, la direction de l'entreprise par une personne, etc. La langue OWL permet l'instanciation d'Object Property pour signaler un fichier. Une connaissance de ce genre, qui nécessite la création d'un nouvel Object Property lors de chaque nouvelle relation, il est intégré au domaine. Bien que cela soit concevable, ce processus peut être lourd en termes de gestion et de cohérence des ontologies.

De plus, le choix peut être fait, dans la conception et la maintenance d'une ontologie, pour prédéterminer la portée des classes et des relations instantanées. Ainsi, la gestion de l'ontologie peut rendre impossible la création d'un grand nombre de Object Property, chacun correspond à une certaine relation entre deux entités [8].

## **2.5 Conclusion**

Dans ce chapitre, nous avons identifié les entités nommées et également nous avons aussi mentionné les différentes formes, les types, les différentes classifications et le rôle des entités nommées. Nous avons par la suite en parlé des ontologies et de ses types ainsi que les domaines d'application. Nous avons également donné une définition du système d'extraction basée ontologie avec une architecture OBIE. En fin on terminé par donner la relation entre les entités nommées et les ontologies. Et ce pour mieux comprendre les chapitres suivants.

Essentiellement les systèmes d'extraction d'informations utilisent des ontologies comme moyen pour décrire formellement les connaissances du domaine exploitées par ces systèmes pour leur fonctionnement. Notre travail consiste à utiliser l'ontologie pour l'extraction et la reconnaissance de différentes entités nommées.

Dans le chapitre suivant, nous présentons des travaux reliés, en proposant et expliquant notre approche avec ses étapes en donnant quelques exemples illustratifs

**Chapitre 03 :**  
Conception du système  
d'extraction d'information à  
base d'ontologie

### 3.1 Introduction

Dans le chapitre précédent nous avons défini l'extraction d'information, et nous avons détaillé ses méthodes et ses tâches avec une architecture générale d'un système d'extraction d'information. Ensuite, nous avons présenté précisément les entités nommées, les formes, les types et les traitements sur les entités nommées, par la suite nous avons en parlé de l'ontologie. Dans ce chapitre, nous présentons une architecture générale de notre approche proposée d'un système d'extraction d'information basée ontologie sur les Catastrophes Naturelles, nous définissons également notre approche ainsi que les étapes de cette proposition.

### 3.2 Travaux reliés

#### 3.2.1 KIM (Knowledge and Information Management)

La plate-forme KIM (Figure 3.1) fournit un cadre de gestion des connaissances et de l'information, des services d'annotation sémantique, d'indexation et de recherche de documents, et enfin une infrastructure d'extraction d'informations. KIM peut accepter et analyser une grande variété de documents types en tant qu'entrée.

L'ontologie utilisée pour le processus OBIE doit également être fournie par l'utilisateur. Les méthodes IE déployées par le système sont basées sur des règles linguistiques et un ensemble de listes contenant des noms d'entités connues sous le nom de listes Gazetteer. Les composants ontologiques qui peuvent être extraites par la plateforme KIM sont des instances et des valeurs de propriété [22].

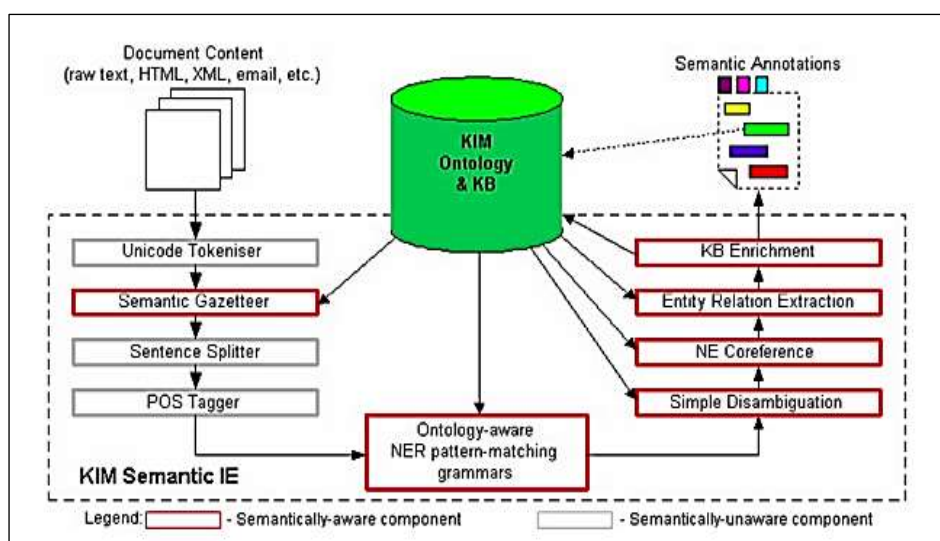


Figure 3.1 : KIM Architecture présentée [23].

### 3.2.2 SPRAT (Semantic Pattern Recognition and Annotation Tool)

SPRAT [24], qui signifie Semantic Pattern Recognition and Annotation Tool, est un autre exemple de système OBIE capable de peupler des ontologies. L'architecture du système est divisée en deux parties.

La première partie est responsable du prétraitement du texte et s'appuie sur les ressources de traitement fournies par le frame work GATE. Le prétraitement est effectué à l'aide de techniques Naturel Langage Processing (NLP) peu profondes telles que la tokenisation, le marquage d'une partie de la parole et l'analyse morphologique. La tokenisation est généralement la première tâche de prétraitement effectuée dans chaque système qui sépare un morceau de texte en unités plus petites appelées jetons. Le marquage de la partie du discours se produit généralement après la tâche de tokenisation et marque un jeton/mot dans un texte avec une partie particulière du discours, en fonction de sa définition et de son contexte. Enfin, l'analyse morphologique est la tâche chargée d'analyser la structure des mots et des parties de mots, tels que les radicaux, les mots racines, les préfixes et les suffixes [25].

La deuxième partie est l'endroit où l'information est extraite. Pour cette raison, les modèles de Hearst ainsi que les modèles lexico-syntaxiques et contextuels sont implémentés en tant que règles JAPE. Les informations extraites de ces modèles combinés sont utilisées pour peupler une ontologie donnée [26].

### 3.2.3 SPEED (Semantics-Based Pipeline for Economic Event Detection)

SPEED [27] qui signifie Semantics-Based Pipeline for Economic Event Detection (**Figure 3.2**) est un Système OBIE proposé en 2013 qui extrait les événements économiques des articles de presse et des mises à jour une base de connaissances en temps réel. Son architecture est basée sur le frame work GATE. Le module de prétraitement du système est similaire à celui utilisé dans SPRAT, le seul ajout étant l'utilisation d'un Word Sense Disambiguator qui identifie le sens d'un mot utilisé dans une phrase et qui est basé sur un algorithme d'Interconnexions Sémantiques Structurelles adapté [28].

Une fois le sens des mots désambiguïsé, le système détecte les événements à l'aide d'un Gazetteer et les valide à l'aide de modèles lexico-sémantiques mis en œuvre comme les règles JAPE. Enfin, les résultats peuvent être utilisés pour mettre à jour la base de connaissances via le module Instanciateur d'ontologie [28].

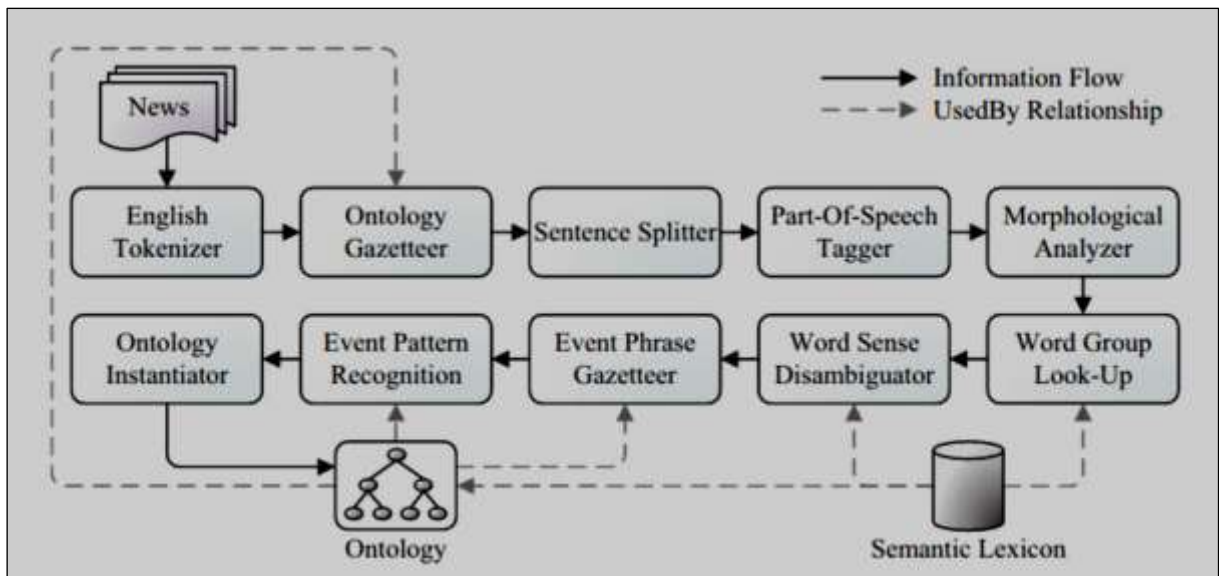


Figure 3.2 : Architecture SPEED telle que présentée [28].

### 3.3 Notre Approche proposée

Les risques et les catastrophes sont de moins en moins tolérés par la société. Ces phénomènes sont désormais pris en compte dans de nombreux domaines et à différentes échelles : celle de l'individu, de la population, du territoire mais aussi celle de la planète.

Quelles que soient les échelles d'analyse, des efforts importants sont mis en œuvre pour atténuer les risques. Des métiers spécifiques ont été créés pour les limiter tant dans le domaine de l'environnement, de l'industrie, de l'agriculture que des services. C'est pourquoi, lorsqu'un accident ou une catastrophe surviennent, la portée de l'évènement a un retentissement important, les sociétés étant de moins en moins aptes à les accepter.

Les catastrophes naturelles sont des événements considérés comme des calamités soudaines qui causent d'énormes dommages à la nature et aux humains, notamment les ouragans, les inondations, les tremblements de terre, les incendies de forêt, les avalanches, les volcans et les tsunamis. Nous avons créé des ontologies dans le domaine de catastrophes naturelles pour extraire les informations qui sont importantes sur des fichiers dans le web.

Dans notre travail, nous étudions les informations et les nouvelles publiées sur Internet pour extraire des informations sur les catastrophes naturelles, leurs caractéristiques, les indicateurs spatiaux et les dommages aux victimes des catastrophes naturelles.

### ❖ Indicateurs spatiaux

Ce sont les informations qui nous aident à connaître et à localiser un lieu dans la publication. Les lieux dans le texte publié dans les journaux électroniques peuvent être des symboles, un dialecte ou une abréviation, de sorte que la tâche d'extraire des informations spatiales est généralement difficile car elle est ambiguë et alias et acronymes au lieu d'utiliser le nom d'origine.

### ❖ Dégats de catastrophe naturelle

Extraction des informations contenues dans le texte publié relatives aux dommages causés par les catastrophes naturelles en termes de pertes humaines et matériels, d'effondrements et d'autres pertes .

### ❖ Caractéristiques de catastrophe naturelle

Ce sont les informations contenues dans le texte relatives aux caractéristiques des catastrophes naturelles, c'est-à-dire la caractéristique qui distingue chaque catastrophe naturelle. Par exemple, la force de l'intensité du tremblement de terre, la température élevée, la quantité de précipitations et d'autres caractéristiques. Il est difficile de les extraire facilement .

## 3.4 Architecture de modèle proposé

L'architecture suivante (**Figure 3.3**) représente la conception détaillée de notre système. L'approche proposée construit les catastrophes naturelles pour extraire les entités nommées et les relations entre elles (la détection des relations).

Le système proposé se compose de plusieurs étapes

- Collection des données.
- Création d'ontologie de domaine.
- Analyse des informations.

**Algorithme.**

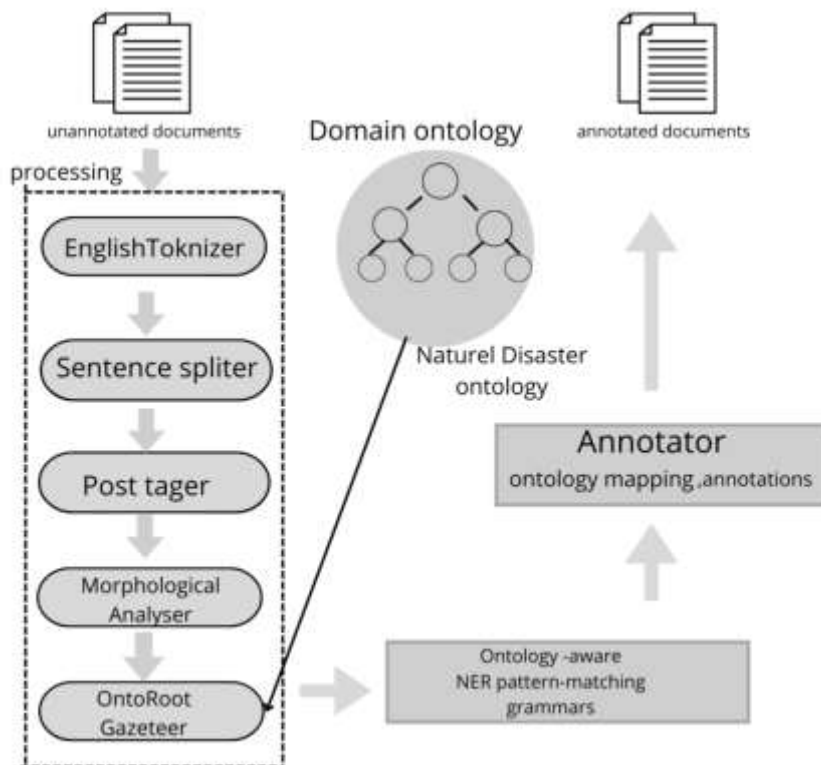


Figure 3.3 : Architecture générale du système proposé.

### 3.5 Etapes de l'approche proposée

#### 3.5.1 Étape de collection des données

Nous obtenons des informations du Web sur Internet. Nous essayons de collecter manuellement des informations sur les événements qui laissent des victimes en raison de catastrophes naturelles. Par exemple, nous avons travaillé sur des textes liés au tremblement de terre, comme le tremblement de terre de Boumerdes .

#### 3.5.2 Étape de création de l'ontologie

La construction de l'ontologie est le thème principal de cette étude, nous avons utilisé une approche descendante dans la construction de l'ontologie. La plupart des concepts abstraits sont d'abord identifiés, puis spécialisés dans des concepts plus spécifiques pour construire notre ontologie NDO Naturel Disaster Ontology de domaine qui représente les connaissances de base dans notre travail. Nous construisons l'ontologie manuellement.

Nous avons développé le contenu de l'ontologie pour le domaine NDO. L'ontologie est implémentée avec l'outil Protégé au format OWL .

Comme nous l'avons précédemment illustré la construction d'ontologies, le développement manuel d'ontologies consiste en les étapes suivantes :

- ✓ Déterminer le domaine et la portée de l'ontologie ;
- ✓ Envisager de réutiliser les ontologies existantes ;
- ✓ Énumérer les termes importants de l'ontologie ;
- ✓ Définir les classes et la hiérarchie des classes ;
- ✓ Définir les propriétés des classes (slots) ;
- ✓ Définir les facettes des fentes ;
- ✓ Créer des instances.

### 3.5.3 Étape d'analyse d'un texte

#### 3.5.3.1 Tokeniser

Le Tokeniser divise le texte en jetons très simples tels que des chiffres, des signes de ponctuation et des mots de différents types. Par exemple, nous distinguons les mots en Majuscule et en Minuscule, et entre certains types de ponctuation, etc. En ajoutant une annotation "Jeton" à chacun, il n'a pas besoin d'être modifié pour différentes applications ou types de texte. L'objectif est de limiter le travail du Tokeniser pour maximiser l'efficacité et permettre une plus grande flexibilité en mettant la charge sur les règles de grammaire, qui sont plus adaptables [29].

#### Les étapes d'analyse des informations :

1)

**Phrase sans analyse**  
Son mémoire est perdu depuis 2 ans, c'est pourquoi il ne connaît pas les gens.

2)

**Décalage des caractères**  
Son mémoire est perdu depuis 2 ans, c'est pourquoi il ne connaît pas les gens.  
0...4.....12.....25.....|.....|.....|.....|.....|



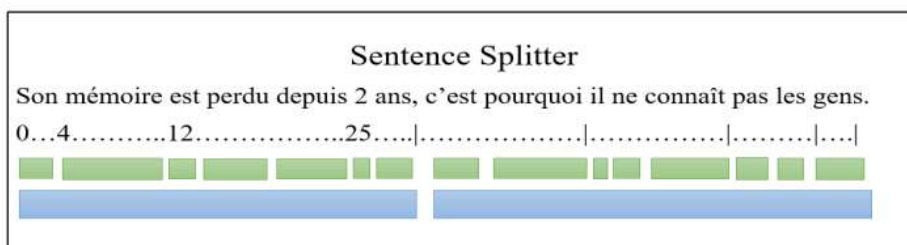
3)



### 3.5.3.2 Sentence Splitter

Le séparateur de phrases est une cascade de transducteurs aux états finis qui segmente le texte en phrases. Le séparateur utilise une liste d'abréviations du répertoire géographique pour aider à distinguer les points de marquage des phrases des autres types. [29] Chaque phrase est annotée avec le type « Phrase ».

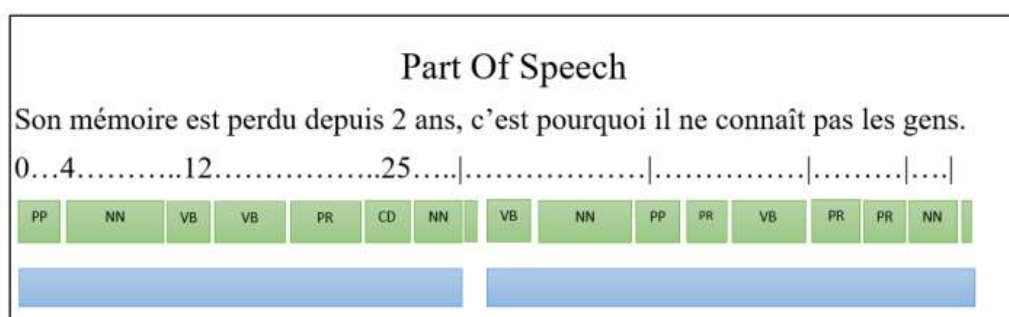
Chaque saut de phrase (comme un point) reçoit également une annotation « Split ». Il a une fonctionnalité « genre » avec deux valeurs possibles : « interne » pour toute combinaison d'exclamation et de point d'interrogation ou un à quatre points et « externe » pour une nouvelle ligne. Le séparateur de phrases est indépendant du domaine et de l'application [29].



### 3.5.3.3 Part Of Speech Tagger

Le tagueur est une version modifiée du marqueur Brill, qui produit une étiquette de partie de discours en tant qu'annotation sur chaque mot ou symbole. Le tagueur utilise un lexique et un ensemble de règles par défaut (résultat d'un apprentissage sur un large corpus tiré du Wall Street Journal). Ces deux éléments peuvent être modifiés manuellement si nécessaire. Deux lexiques supplémentaires existent ; un pour les textes tout en majuscules et un pour les textes tout en minuscules. Pour les utiliser, le lexique par défaut doit être remplacé par le lexique approprié au moment du chargement. L'ensemble de règles par défaut doit toujours être utilisé dans ce cas. [29] Le tagueur ANNIE Part-of-Speech nécessite certains paramètres :

- Encoding
- Lexicon URL
- Rules URL
- Document
- Input AS Name
- Output AS Name
- Base Token Annotation Type
- Base Sentence Annotation Type
- Output Annotation Type
- Pos Tag All Tokens
- Fail On Missing Input Annotations [29]



### 3.5.3.4 Analyseur Morphologique

Un analyseur morphologique est le module qui réduit les morphes à leur forme canonique.

Chaque mot peut avoir une variété de formes et chaque forme peut décrire le même concept, mais d'un point de vue différent.

Par exemple, le nom « maladie » est une transformation du nom « maladies » qui décrit le concept d'une maladie au pluriel. L'analyseur morphologique dans ce cas recevra le jeton qui a été enrichi avec une partie de l'information de langue et créera deux nouvelles fonctions d'annotation "root=disease" et "suffix=s". Il s'agit d'une composante importante du pipeline puisque le Répertoire topographique Onto Root est fortement basé sur la racine caractéristique des jetons à rechercher dans l'ontologie [29].

### 3.5.3.5 Onto Root Gazetteer

Onto Root Gazetteer (**Figure 3.4**) est un type de répertoire géographique créé dynamiquement qui est, en combinaison avec quelques autres ressources GATE génériques, capable de produire des annotations basées sur l'ontologie sur le contenu donné

en ce qui concerne l'ontologie donnée. Ce répertoire fait partie de Plugin 'Gazetteer Ontology Based' qui a été développé dans le cadre du projet TAO.

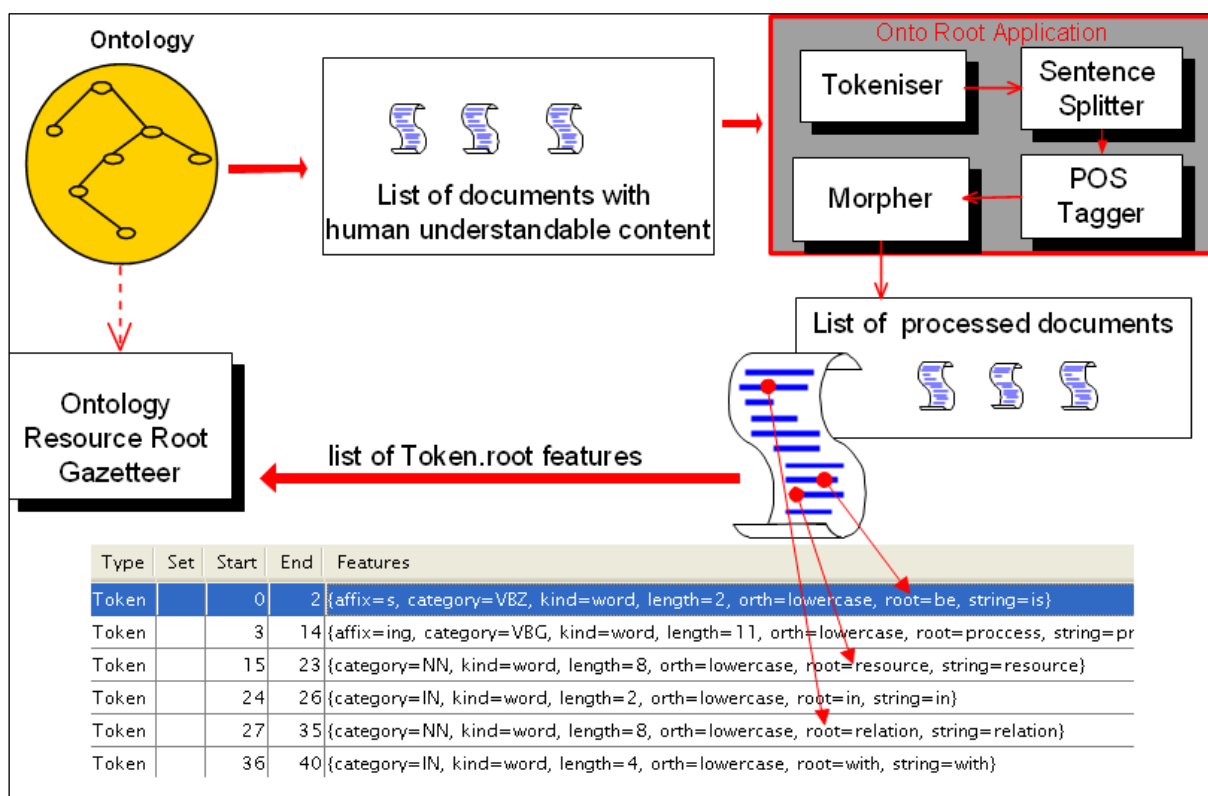


Figure 3.4 : Building Ontology Resource Root (Onto Root) Gazetteer from the Ontology

### 3.5.3.6 Semantic Tagger

Le marqueur sémantique d'ANNIE (A Nearly-New Information Extraction System) est basé sur le langage JAPE (Java Annotation Patterns Engine). Il contient des règles qui agissent sur des annotations, fonctionnalités et valeurs attribuées dans les phrases précédentes. Il contient aussi des règles qui agissent sur des annotations, fonctionnalités et valeurs qui doivent être attribuées manuellement [23]. Les types d'annotations, les caractéristiques et les valeurs possibles par défaut produits par ANNIE sont basés sur les types d'entités MUC d'origine et sont les suivants :

- Personne : genre : masculin, féminin ;
- Localisation : loc Type, région, aéroport, ville, pays, département, province, autre ;
- Organisation : orgType : entreprise, département, gouvernement, journal, équipe, autre ;
- Argent ;
- Pourcentage ;
- Date : genre : date, heure, date Heure;

- Adresse : type : email, URL, téléphone, code postal, complet, ip, autre ;
- Identifiant ;
- Inconnu. [29]

### 3.5.4 Algorithme

JAPE est un moteur de modèle d'annotation Java. JAPE fournit une transduction aux états finis sur annotations basées sur des expressions régulières. JAPE est une version de CPSL (Common Pattern Specification Language). JAPE permet de reconnaître les expressions régulières dans les annotations sur les documents : un langage régulier ne peut décrire que des ensembles de chaînes de caractères, pas des graphes, et le modèle d'annotations de GATE est basé sur des graphes.

Typiquement, les expressions régulières sont appliquées à des chaînes de caractères, une simple séquence linéaire d'éléments. Le résultat est que, dans certains cas, le processus de correspondance n'est pas déterministe (c'est-à-dire que les résultats dépendent de facteurs aléatoires tels que les adresses auxquelles les expressions régulières sont utilisées et de facteurs aléatoires tels que les adresses auxquelles les données sont stockées dans la machine virtuelle).

Lorsqu'il existe une structure dans le graphe à mettre en correspondance et qui nécessite plus que la puissance d'un automate régulier, JAPE choisit arbitrairement une alternative. Cependant, il s'avère que dans de nombreux cas utiles les données stockées dans les graphes d'annotation dans GATE (et d'autres systèmes de traitement du langage) peuvent être considérées comme séquences simples, et correspondre de manière déterministe avec des expressions régulières.

Ci-dessous un exemple de règle JAPE :

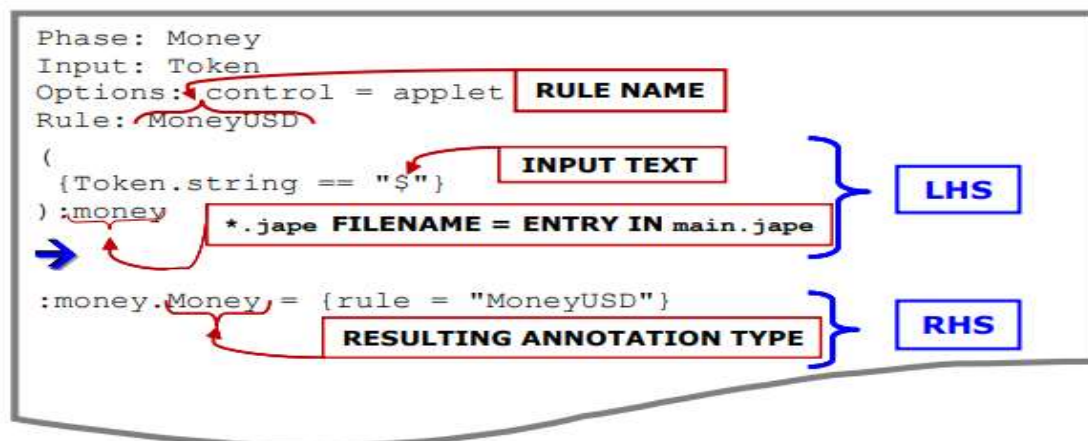


Figure3.5: Exemple de règle avec le formalisme JAPE

### **3.6 Conclusion**

Dans ce chapitre nous avons détaillé les étapes de notre approche d'extraction basée ontologie, nous avons expliquée également les différentes étapes qui sont la collecte des données, création d'ontologie, l'étape d'analyse et l'algorithme utilisé avec des exemples pour plus d'éclaircissement.

Dans le chapitre suivant nous allons implémenter notre modèle de système d'extraction a base d'ontologie et d'autre techniques sur des documents en ligne.

# Chapitre 04 :

# Implémentation

### 4.1. Introduction

Dans ce chapitre nous allons appliquer les étapes de notre approche proposée dans le chapitre précédent, conception d'un système d'extraction des entités nommés basée ontologie, puis nous allons choisir l'outil Protégé pour créer l'ontologie et aussi la plate forme GATE développer comme un environnement d'application de notre travail.

### 4.2. Protégé

Pour la construction d'ontologies, nous avons utilisé Protégé 4 [30] qui est un éditeur d'ontologies de base de connaissances fournissant une interface utilisateur graphique. Il est choisi pour notre construction d'ontologies car il offre une meilleure flexibilité pour la méta-modélisation, permet la construction d'ontologies de domaine et personnaliser les formulaires pour saisir les données. Il est généralement destiné à l'ingénierie des connaissances et à la modélisation conceptuelle [31]. La figure 4.1 montre une capture d'écran de Protégé qui inclut la hiérarchie des classes et son graphique correspondant.

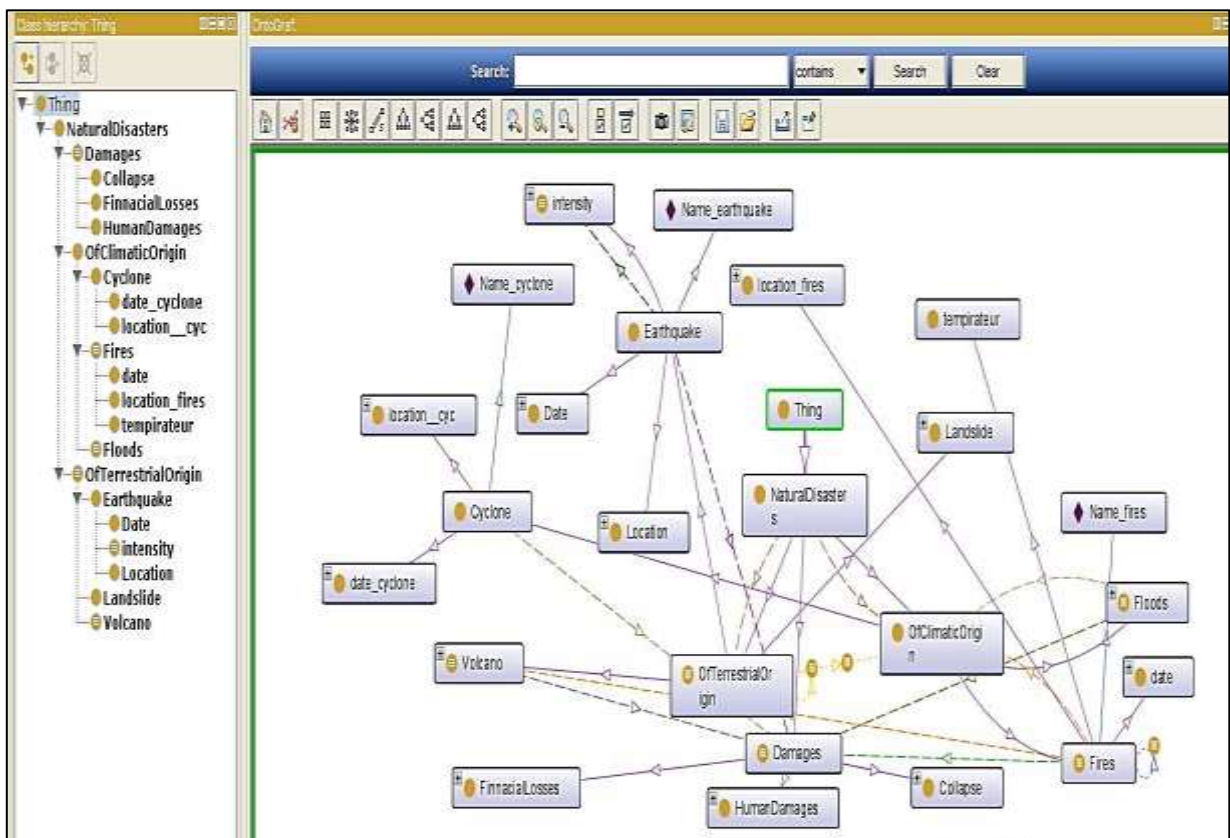


Figure 4.1: Protégé Screenshot .

### 4.2.1 Création de l'ontologie

Comme nous l'avons précédemment illustré la construction d'ontologies, le développement manuel d'ontologies consiste selon les étapes suivantes :

#### **Étape 1 : Déterminer le domaine et la portée de l'ontologie**

La première étape du développement de l'ontologie est la définition du domaine et de la portée de l'ontologie. Pour déterminer le domaine et la portée de l'ontologie, nous devons savoir quel domaine l'ontologie couvrira, le but de l'ontologie et le type de question si les informations contenues dans l'ontologie peuvent fournir des réponses, et qui utilisera l'ontologie.

En répondant à ces questions, nous pouvons dire que le domaine de l'ontologie couvrira les domaines de l'actualité, ce qui est utilisé dans notre approche de l'annotation.

#### **Étape 2 : Envisager de réutiliser les ontologies existantes**

Cette étape consiste à vérifier s'il existe une ontologie développée précédemment dans le même domaine. Si une telle ontologie existe, il est plus facile de modifier l'ontologie existante pour l'adapter à ses besoins que d'en créer une nouvelle. Et parce que nous n'avons pas trouvé d'ontologie précédemment créée sur le domaine des Catastrophes Natural NDO (Naturel Disaster Ontology), nous avons négligé cette étape.

#### **Étape 3 : Énumérer les termes importants dans l'ontologie**

Cette étape peut être considérée comme une activité de remue-méninges, dans laquelle nous listons les mots que nous voulons utiliser, pour démontrer les termes de l'ontologie et les propriétés qui peuvent avoir.

Nous avons également bénéficié des documents collectés pour acquérir les connaissances à propos des termes de catastrophes naturelle.

#### **Étape 4 : Définir les classes et la hiérarchie des classes**

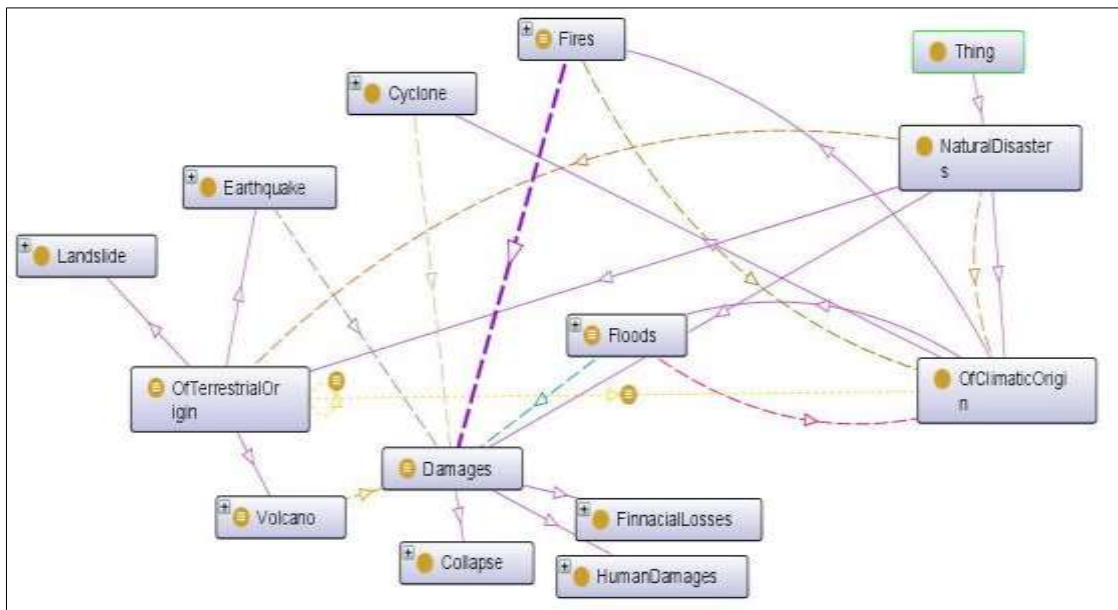
Cette étape définit les classes (concepts) utilisées dans notre domaine ontologique. Nous définissons des classes et sous-classes liées à notre domaine. Le tableau 4.1 décrit l'ontologie des classes.



Classes/Sub classes	In English	Classes/Sub classes	In French
<b>Damages</b>		<b>Dégats</b>	
<b>Collapse</b>		Effondrement	
<b>Financial losses</b>		Pertes financières	
<b>Human Damages</b>		Dégâts humains	
<b>Of climatic origin</b>		<b>D'origine climatique</b>	
<b>Cyclone</b>		Cyclone	
<b>Fires</b>		Les feux	
<b>Floods</b>		Inondations	
<b>Of terrestrial origin</b>		<b>D'origine terrestre</b>	
<b>Earthquake</b>		Tremblement de Terre	
<b>Volcano</b>		Volcan	
<b>Landslide</b>		Glissement de terrain	

**Table 4.1:** Ontology Classes and Sub-classes

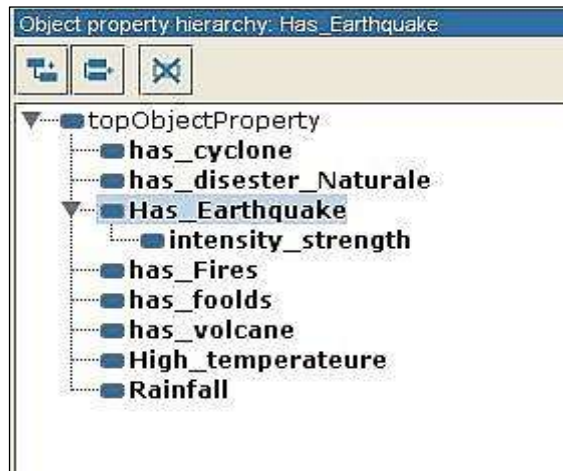
La figure 4.2 représente les classes d'ontologie de niveau supérieur. La chose, qui représente la classe de toutes choses. La classe (DNO) qui est la classe racine, et d'autres sont les sous-classes.



**Figure 4.2:** Classes d'ontologie de haut niveau

**Étape 5 : Définir les propriétés des classes (slots).**

Définissez les propriétés d'objet (relations) entre les classes, dont le rôle est de connecter les concepts de l'ontologie.



**Figure 4.3** : Hiérarchie de propriété d'objet.

**Exemple** : “Has \_Earthquake”: le domaine de cette propriété est “Earthquake class” et la gamme est (the range) “Damages class”.

**Étape 6 : Définir les facettes des fentes**

Les emplacements (parfois appelés rôles ou propriétés) ont différentes facettes (parfois appelées restrictions de rôle) qui décrivent le type de valeur, les valeurs autorisées, le nombre de valeurs (cardinalité) et d'autres caractéristiques des valeurs que l'emplacement peut prendre. Par exemple, nous avons une classe associée à une relation qui a une valeur.

**Étape 7 : Créer des instances**

La création d'instances (individus) est une étape très importante pour enrichir l'ontologie est qui est en relation directe avec les classes et sous-classes.

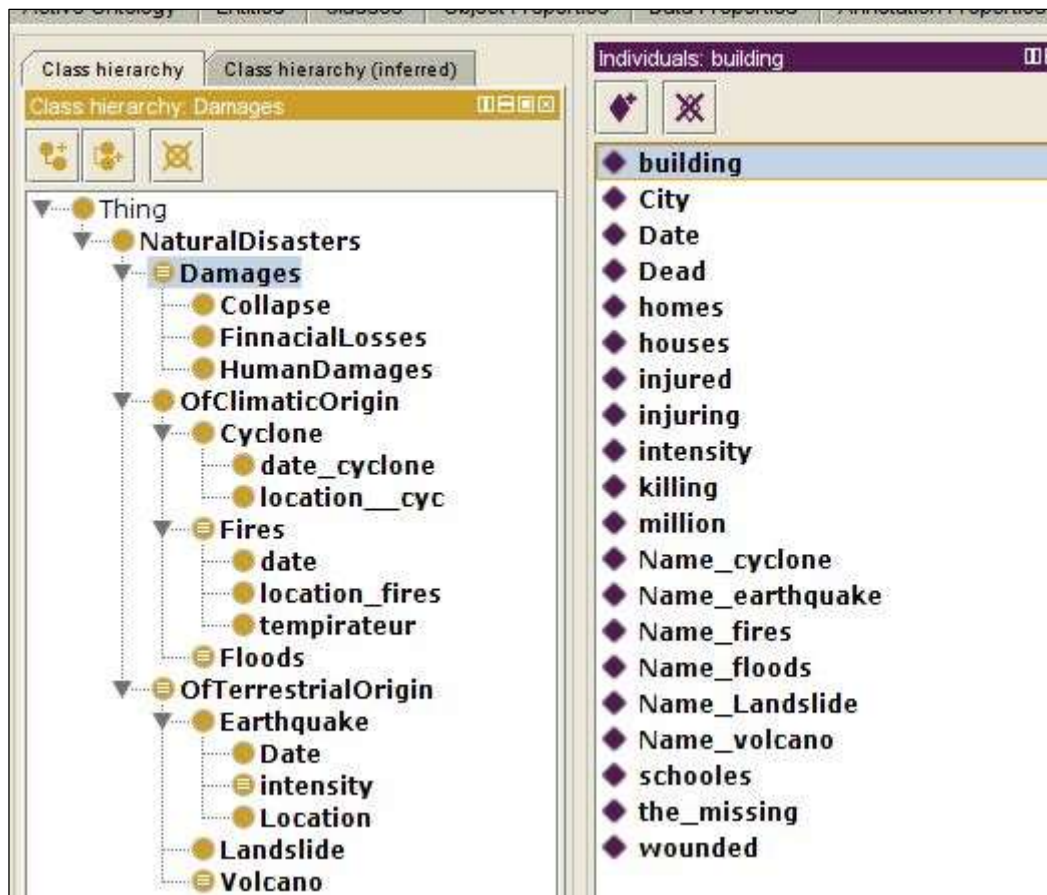


Figure 4.4 Liste de certaines instances d'ontologie

### 4.3. GATE

GATE [32] (General Architecture for Text Engineering) est un logiciel source développé par l'Université de Sheffield. C'est une infrastructure pour développer et déployer des composants logiciels qui traitent le langage humain. GATE est un outil efficace utilisé pour effectuer certains NLP (Natural Language Processing), il possède de nombreuses fonctionnalités, telles que l'annotation manuelle, l'annotation automatique, l'utilisation d'une variété de gazetteer, la récupération d'informations et le traitement basé sur l'ontologie [33].

GATE comprend un système d'extraction d'informations appelé ANNIE (A Nearly-New Information Extraction System) qui est un ensemble de modules comprenant un tokenizer, un gazetteer, un séparateur de phrases, un étiqueteur de partie de parole, un transducteur d'entités nommées et un étiqueteur de coréférence.

ANNIE peut être utilisé tel quel pour fournir une fonctionnalité d'extraction d'informations de base ou fournir un point de départ pour des tâches plus spécifiques.

GATE comprend différents composants que nous pouvons identifier ses composants qui se déclinent en trois types :

➤ **Ressources langage** (LRs : Language Resources)

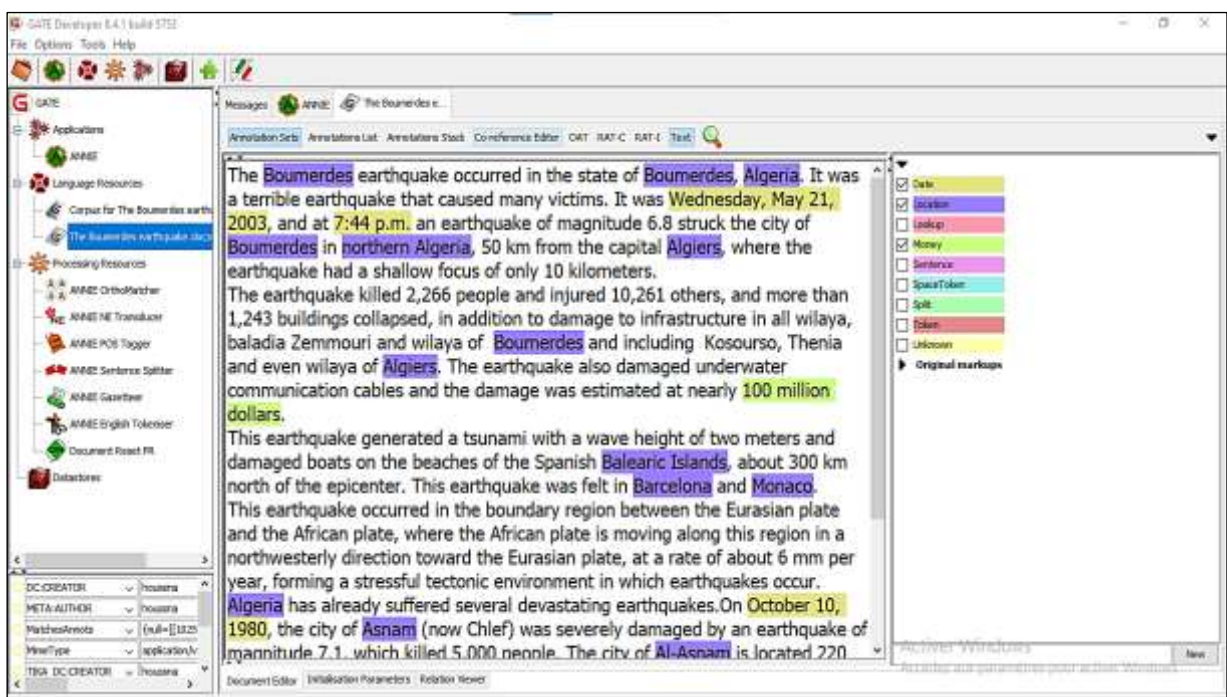
Il s'agit d'un certain nombre de données linguistiques tels que les documents, les corpus et les ontologies.

➤ **Ressources de traitement** (Algorithmique) (PRs : Processing Resources)

Ce sont des programmes ou des algorithmes qui feront une sorte de traitement sur le texte, c'est-à-dire Tokenisation ou dictionnaire de recherche, analyse, etc. Dans la majorité des cas, les PR sont utilisés pour traiter les données fournies par les LR.

➤ **Ressources de visualisation** (VRs : Visual Resources)

Ce sont des composants pour l'utilisateur graphique interface et permet de visualiser et de modifier d'autres types de ressources (Figure 4.5).



**Figure 4.5** : Exemple de l'interface dans GATE

### 4.3.1 ANNIE

ANNIE (A Nearly-New Information Extraction System), est un composant de GATE, formé de plusieurs modules parmi lesquels un analyseur lexical, un Gazetteer, un segmenteur de phrases, un étiqueteur, un module d'extraction d'entités nommées et un module de détection de coréférences. ANNIE offre toute la gamme de Processing Ressources nécessaires au dépistage d'information sur les textes (Information Extraction). Il offre également entre autres les outils pour le traitement de phrases, pour la détection des entités et pour la détection de références entre les sections d'un texte.

### 4.3.2 Le formalisme JAPE

Une partie des différents modules proposés dans GATE est basée sur JAPE (Java Annotation Patterns Engine), un transducteur aux états finis permettant de reconnaître des expressions régulières sur les annotations. Ce système s'avère très utile en extraction d'informations car il permet de définir les contextes d'apparition des éléments à extraire pour ensuite les repérer et les annoter.

Le principe est de combiner différentes annotations « basiques » (tokens, syntagmes, relations syntaxiques, etc.) pour en créer de nouvelles plus complexes (entités nommées, relations, évènements, etc.) : cela revient à l'écriture de règles de production et donc à l'élaboration d'une grammaire régulière. Une grammaire JAPE se décompose en plusieurs phases exécutées consécutivement et formant une cascade d'automates aux états finis. Chaque phase correspond à un fichier «.jape » et peut être constituée d'une ou plusieurs règle(s) écrite(s) selon le formalisme associé à JAPE.

Classiquement, ces règles sont divisées en deux blocs : une partie gauche (« Left Hand Side » ou LHS) définissant un motif d'annotations à repérer et une partie droite (« Right Hand Side » ou RHS) contenant les opérations à effectuer sur ce motif. Le lien entre ces deux parties se fait par l'attribution d'une étiquette au motif (ou à ses constituants) en LHS et par sa réutilisation en RHS pour y appliquer les opérations nécessaires.

### 4.3.3 OntoRootgazeteer

Le répertoire toponymique OntoRoot est un type de répertoire toponymique créé dynamiquement qui, en combinaison avec quelques autres ressources génériques GATE, est capable de produire des annotations basées sur l'ontologie et sur le contenu donné par rapport à l'ontologie donnée.

OntoRoot Gazetteer relie le texte à une ontologie en créant des annotations Lookup qui proviennent de l'ontologie plutôt que d'un répertoire toponymique par défaut.

Le répertoire toponymique OntoRoot a besoin de quelques paramètres obligatoires pour être initialisé comme le montre la Figure 4.6 et qui sont :

- Ontologie LR.
- Application Root Finder

Name	Type	Required	Value
caseSensitive	Boolean	✓	true
considerHeuristicRules	Boolean	✓	false
considerProperties	Boolean	✓	true
ontology	Ontology	✓	OND1_00008
propertiesToExclude	String		
propertiesToInclude	String		
rootFinderApplication	CorpusController	✓	rootfinder

**Figure 4.6 :** Paramètres de l'OntoRootgazeteer

L'application de rootfinder se compose des ressources de traitement suivantes (PR), illustrées dans la figure 4.7 et dont les états sont les suivants :

- PR de réinitialisation du document : permet de rétablir l'état initial du document en supprimant tous les ensembles d'annotations et leur contenu et l'état d'origine, en supprimant tous les ensembles d'annotations et leur contenu ;
- Tokenizer anglais : qui tokenise le texte avec des types de tokens tels que mot, nombre, ponctuation et espace ;
- Regex Sentence Splitter : qui segmente le texte en phrases ;
- ANNIE POS Tagger : Part of Speech Tagger : qui produit une étiquette de partie de la parole comme une annotation sur chaque mot ou symbole ;
- Analyseur morphologique GATE : qui trouve les valeurs de la racine et de l'afixe d'un token et les ajoute comme caractéristiques aux tokens ;



I	Name	Type
	Document Reset PR 0000A	Document Reset PR
	ANNIE English Tokeniser 0000B	ANNIE English Tokeniser
	RegEx Sentence Splitter 0000E	RegEx Sentence Splitter
	ANNIE POS Tagger 0000F	ANNIE POS Tagger
	GATE Morphological analyser 00010	GATE Morphological analyser

**Figure 4.7:** Les PRs de l'application rootfinder

#### 4.4. Différentes méthodes appliquées

Les méthodes d'extraction détaillées ci-dessus sont implémentées dans l'environnement de traitement de texte GATE. Nos chaînes d'extraction de relations sont utilisées comme base pour les unités d'extraction des entités individuelles de la bibliothèque d'ontologies.

Nous avons fait une comparaison entre trois méthodes pour l'extraction des entités nommées qui sont :

##### 4.4.1 Première méthode appliquée (par défaut)

Dans cette méthode on traite les documents avec l'application par défaut ANNIE, le système extrait les résultats suivants :

- **Entité Nommée Spatiale (location)**

On exécute un document d'une catastrophe naturelle et le système détecte les différentes locations suivantes :

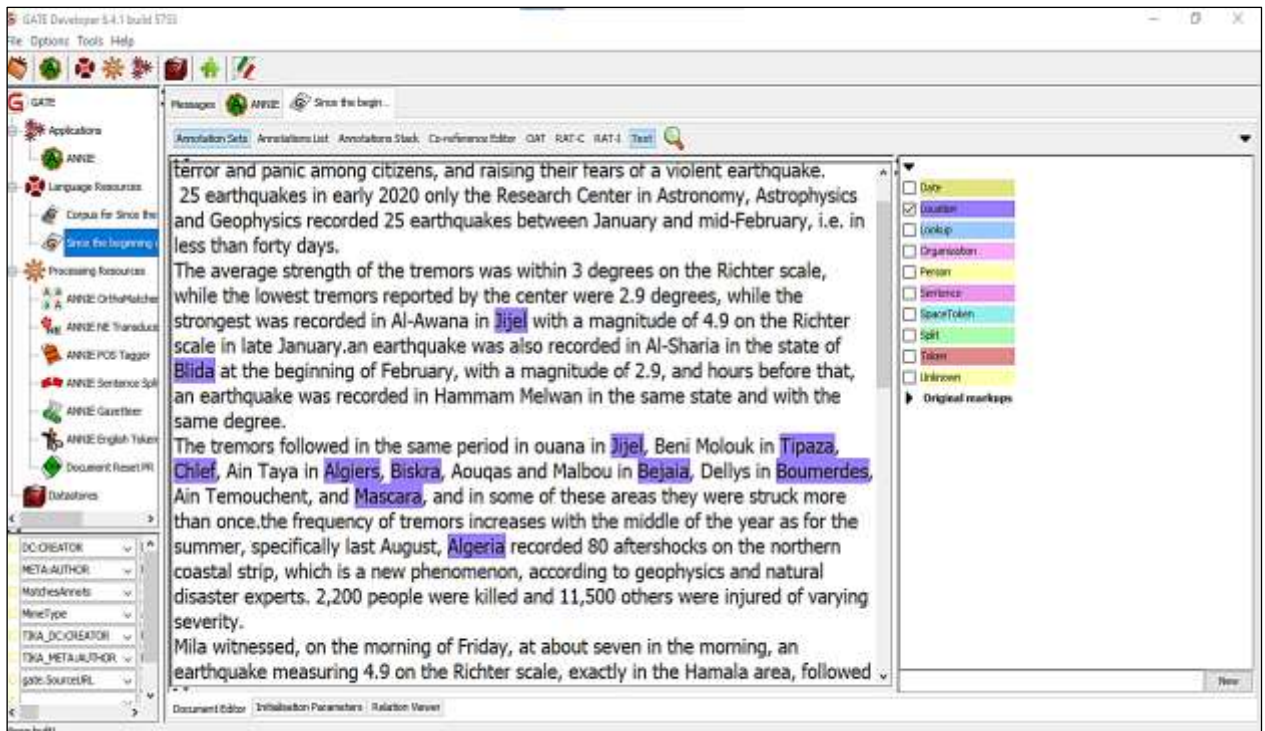


Figure 4.8 : Exécution d'entités nommées spatiales avant l'ajout des règles JAPE

- **Entité Nommée Caractéristiques des catastrophes naturelles**

On exécute un document avant l'utilisation des règles JAPE et le système détecte aucune entité nommée « caractéristiques » de catastrophe naturelle, le résultat c'était les suivant :

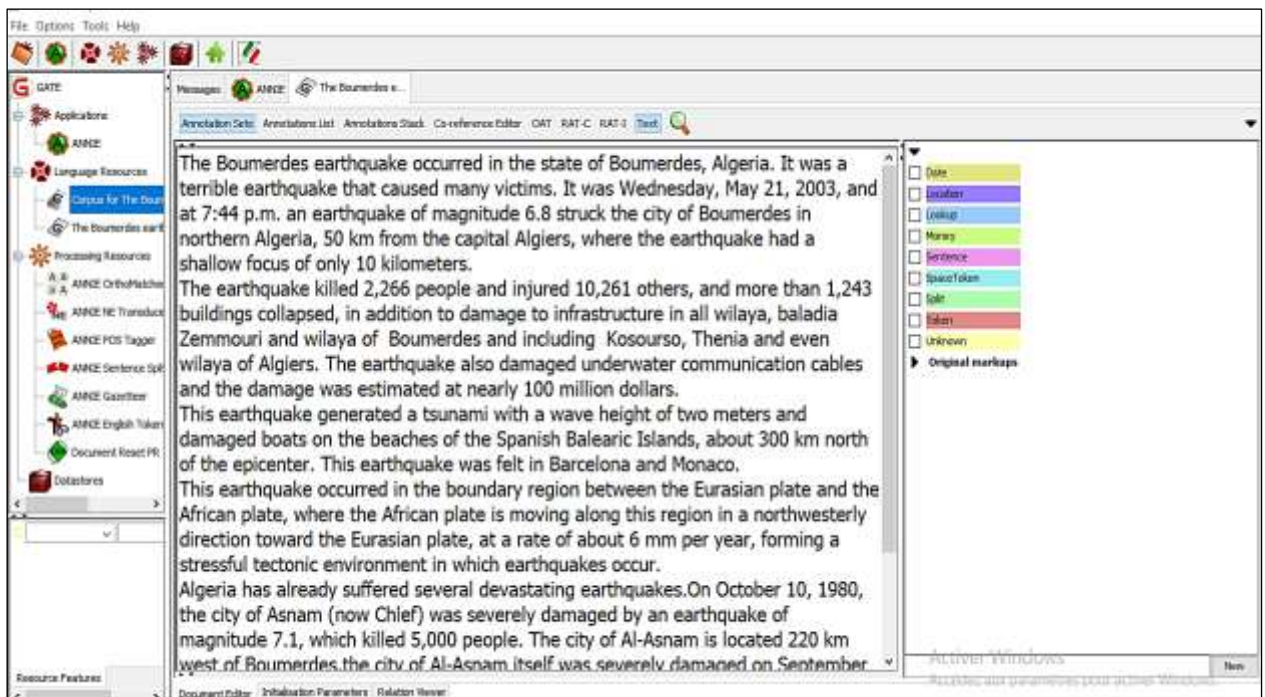
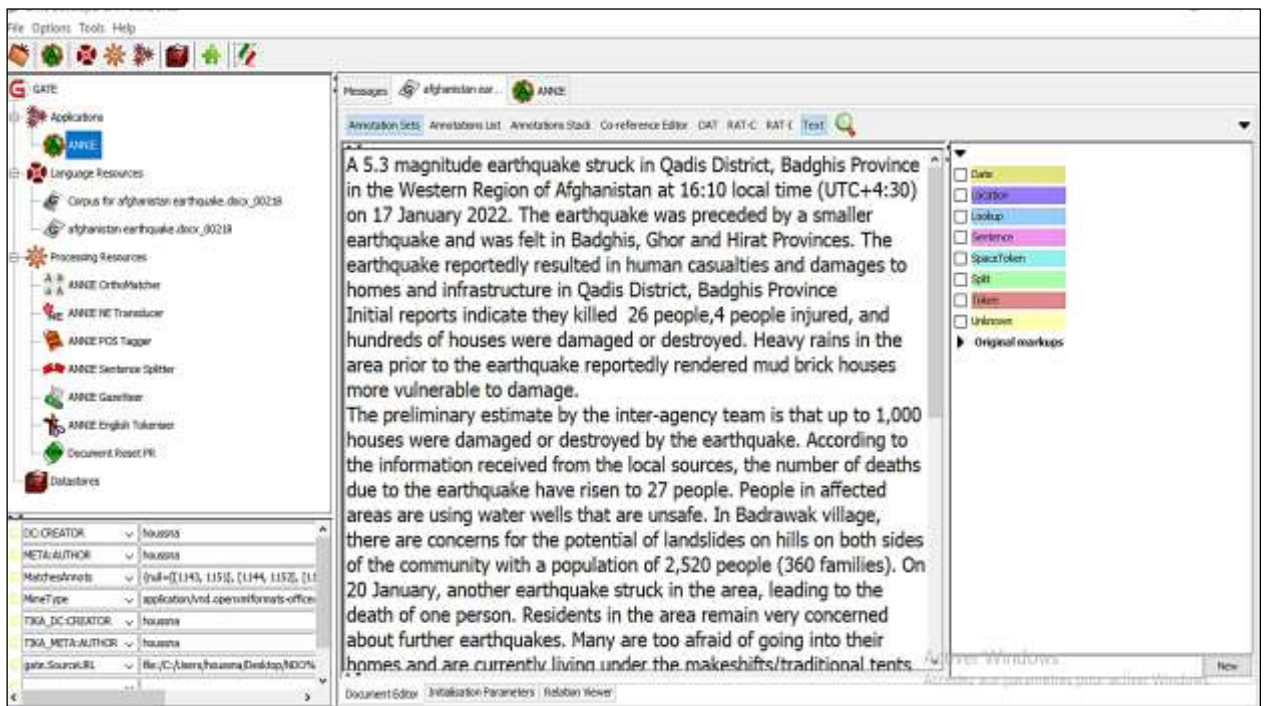


Figure 4.9 : Exécution d'entités nommées (caractéristiques) avant l'ajout des règles JAPE



- **Entité Nommée Dégâts des catastrophes naturelles**

On Exécute un document avant l'utilisation des règle JAPE et le système détecte aucun entité nommée «Dégâts » de catastrophe naturelle, le résultat c'était les suivant :



**Figure 4.10:** Exécution d'entités nommées (Dégâts) avant l'ajout des règles JAPE.

#### 4.4.2 Deuxième méthode appliquée (basée sur les règles JAPE)

Dans cette méthode nous avons créé des règles JAPE qui vont aider le système à identifier les différentes entités nommées contenues dans les documents que nous avons traité, et ceci un exemple de règle que nous avons utilisé :

```

phase: locationcityofdisaster
Input: Token Lookup
Options: control = appelt

Rule: locationcity
(
  {Lookup.majorType == city}

  (
    {Lookup.majorType == city}
  )?
)
:ci
-->
:ci.location={Rule="locationcity"}

```

- **Entité Nommée Spatiale (location)**

On Exécute le document après l'utilisation des règle JAPE et le système détecte plus des entités nommées «Location » de catastrophe naturelle ,le résultat c'était les suivant :

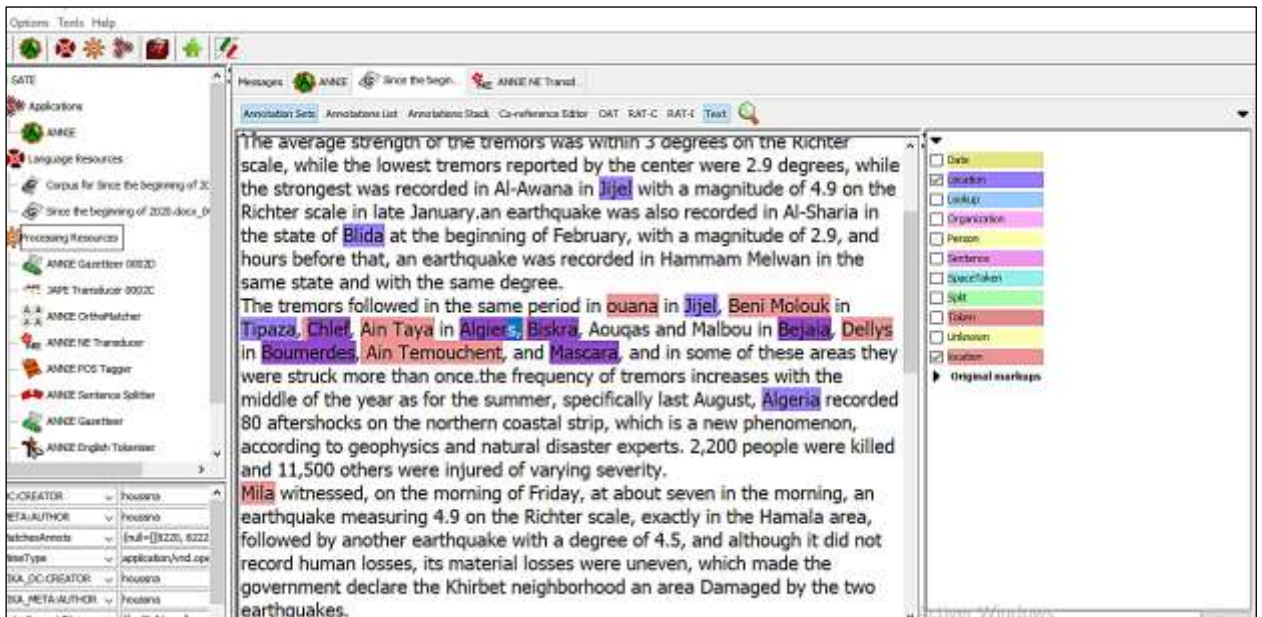


Figure 4.11: Exécution d'entités nommées spatiales après l'ajout des règles JAPE

- **Entité Nommée Caractéristiques des catastrophes naturelles**

On Exécute un document après l'utilisation des règle JAPE et le système maintenant détecte les entité nommée «caractéristiques » de catastrophe naturelle ,le résultat c'était les suivant

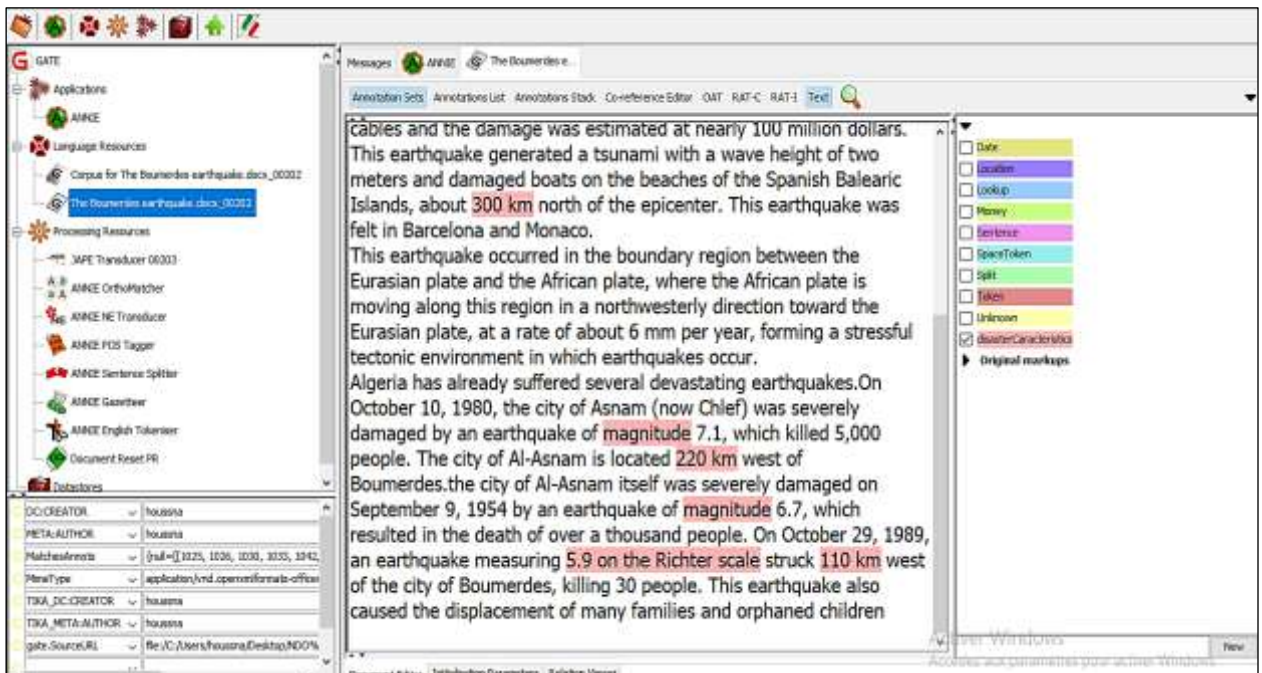
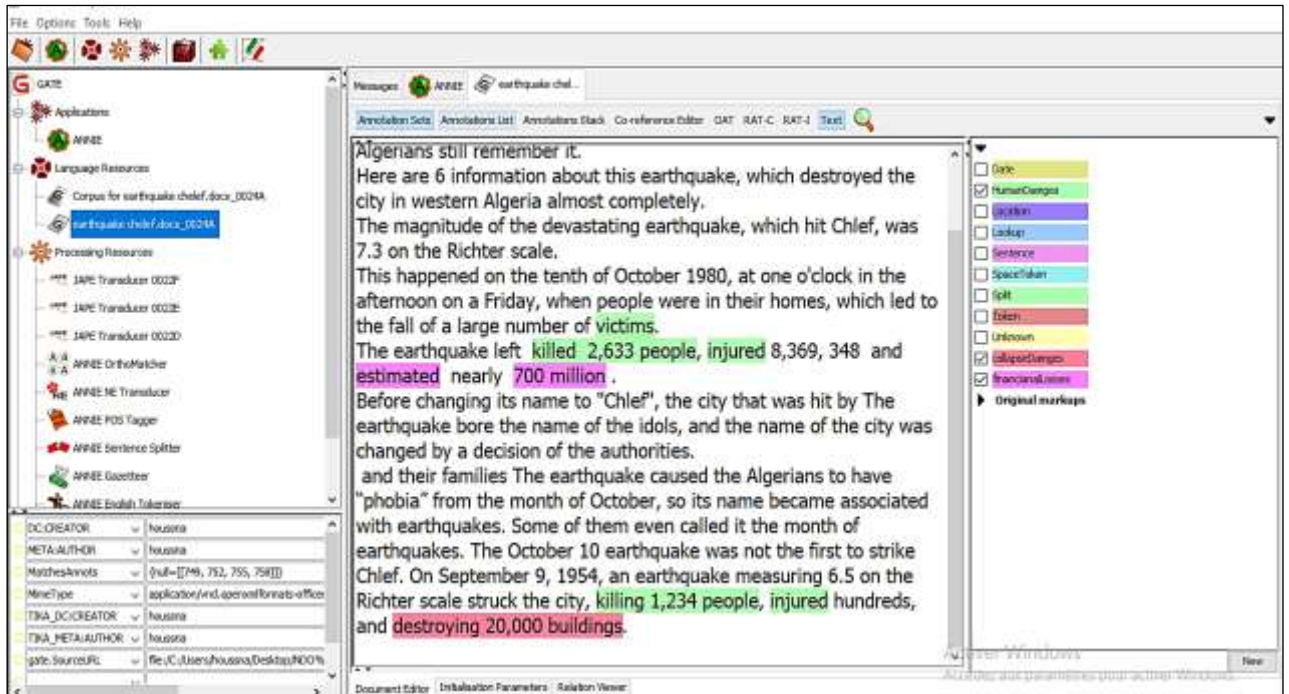


Figure 4.12: Exécution d'entités nommées (caractéristiques) après l'ajout des règles JAPE

- **Entité Nommée Dégâts des catastrophes naturelles**

On Exécute un document après l'utilisation des règle JAPE et le système détecte plus d'entité nommée «Dégâts » de catastrophe naturelle, le résultat c'était les suivant :



**Figure 4.13** : Exécution d'entités nommées (Dégâts) après l'ajout des règles JAPE

#### 4.4.3 Troisième méthode appliquée (basée ontologie)

Dans la dernière méthode nous utilisons l'ontologie pour connaitre les différentes entités nommées contenues dans les textes traités. Le système va chercher dans notre ontologie de domaine et affiche les résultats comme suit :

- **Entité Nommée Spatiale (location)**

On Exécute le document par l'ontologie et le système détecte plus des entités nommées «Location » de catastrophe naturelle ,le résultat c'était les suivant :



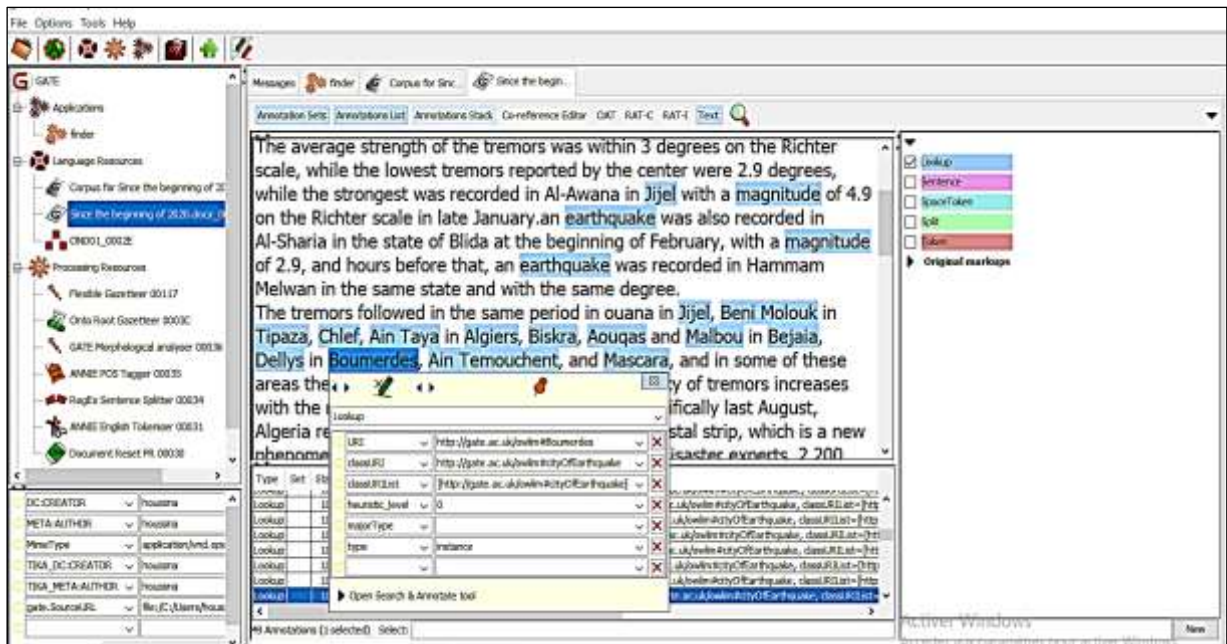


Figure 4.14 : Exécution d'ontologie (lookup)(location).

- Entité Nommée Caractéristiques des catastrophes naturelles :

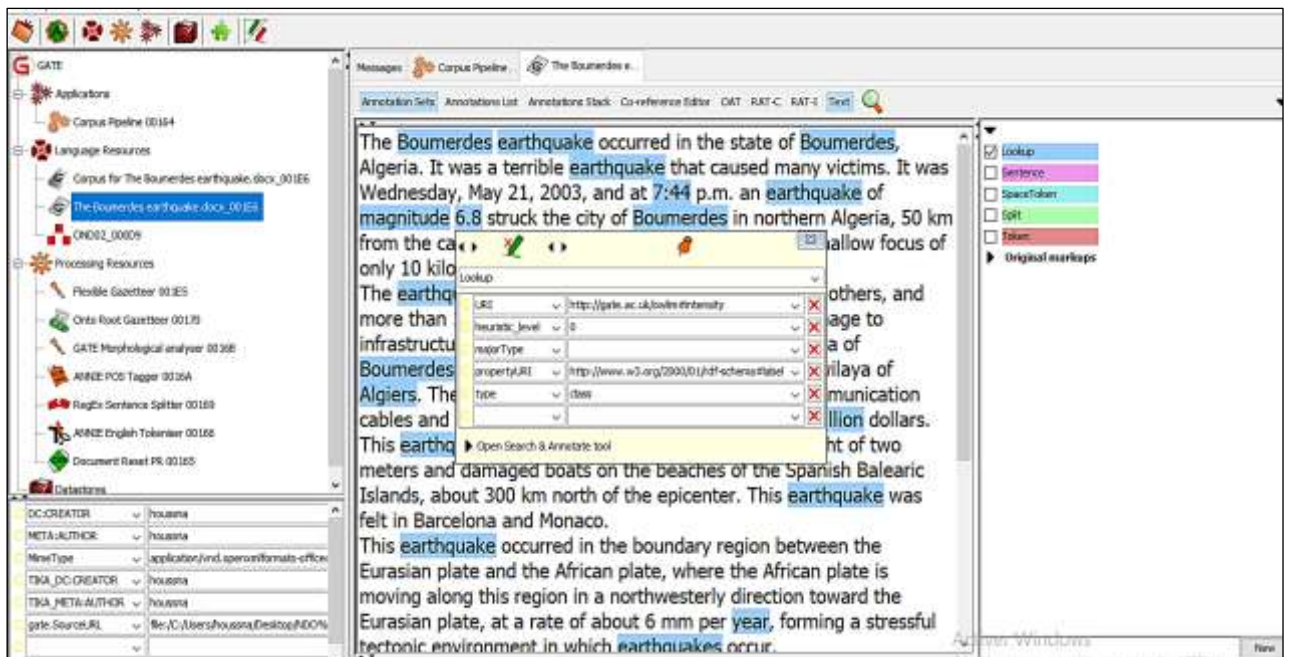


Figure 4.15: Exécution avec l'ontologie (lookup caractéristiques).

. Entité Nommée Dégâts des catastrophes naturelles

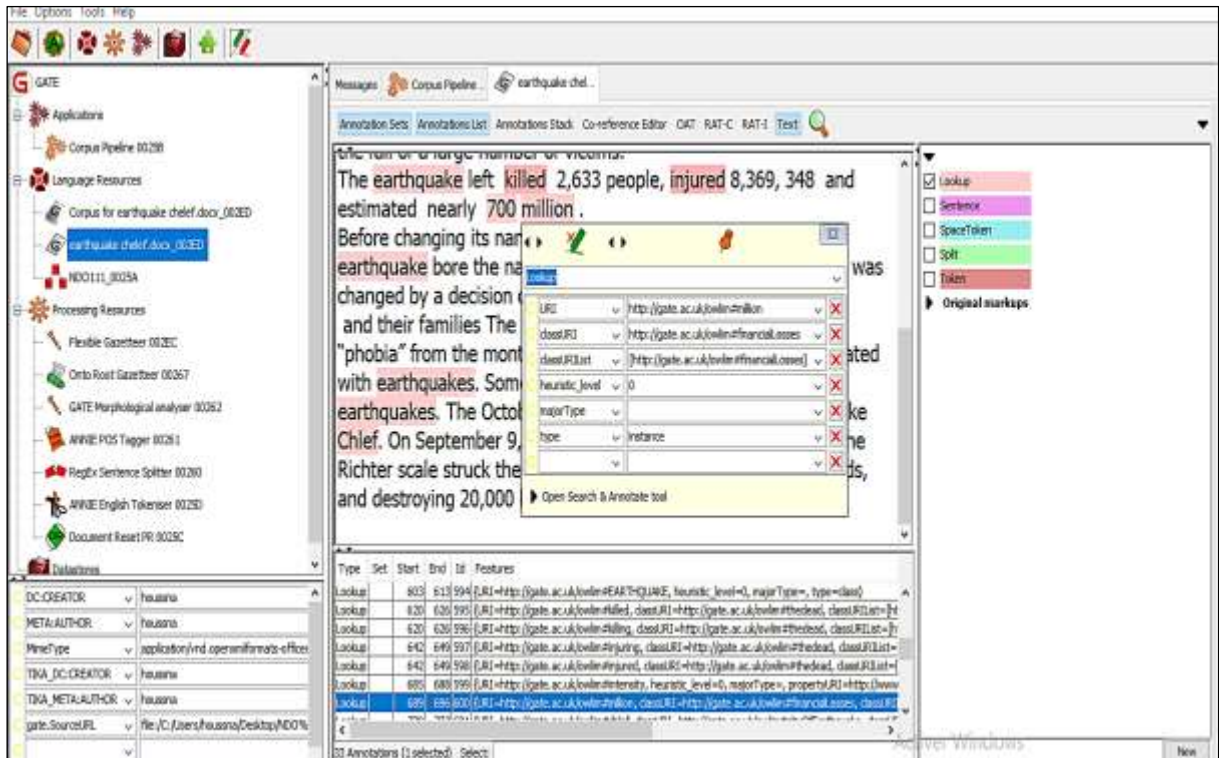


Figure 4.16 : Exécution a base d'ontologie (lookup).

4.5 Evaluation

4.5.1 Métriques d'évaluation des entités nommées

- ❖ Le rappel, la précision et la F-mesure sont des mesures largement utilisées dans les évaluations en TALN ;
- ❖ La précision est le pourcentage des résultats corrects parmi les résultats obtenus ;
- ❖ Le rappel est le pourcentage des résultats corrects parmi les résultats qu'on doit trouver ;
- ❖ Pour le domaine de l'extraction des ENS, les taux de la précision et du rappel sont calculés selon les formules suivantes :

$$\text{Précision} = \frac{\text{Nombre d'ENS correctemnt reconnues}}{\text{Nombre d'ENS reconnues}}$$

$$\text{Rappel} = \frac{\text{Nombre d'ENS correctemnt reconnues}}{\text{Nombre d'ENS dans le corpus}}$$

La F-mesure est la combinaison de la précision et du rappel et leur pondération. La formule de la F-mesure est [12] :

$$F_{\text{mesure}} = \frac{2(\text{précision} * \text{rappel})}{\text{précision} + \text{rappel}}$$

Nous avons présenté ci-dessous les tables d'évaluation des entités nommées :

Liste des entités nommées extraites des dix documents que nous avons traités.

**Table 4.2 :** Numéros d'entité nommés extraits de documents.

les entités nommées	Système par défaut	système abas de règle Notre approche 1	système Abas d'ontologie noter Approche 2	les entités nommées détecte+ non détecte
Location	67	73	74	77
caractéristique	0	39	41	43
catastrophe naturelle Dégât	0	23	24	27

**Table 4.a :** Evaluation : Entité Nommée Spatiale

Catégorie	Précision	Rappel	F-mesure
Méthode 1	<b>0.28</b>	<b>0.87</b>	<b>0.42</b>
Méthode 2	<b>0.71</b>	<b>0.94</b>	<b>0.81</b>
Méthode 3	<b>0.85</b>	<b>0.96</b>	<b>0.90</b>

**Table 4.b :** Evaluation : Entité Nommée Dégâts des catastrophes naturelles

Catégorie	Précision	Rappel	F-mesure
Méthode 1	<b>0</b>	<b>0</b>	<b>0</b>
Méthode 2	<b>0.2</b>	<b>0.85</b>	<b>0.32</b>
Méthode 3	<b>0.80</b>	<b>0.88</b>	<b>0.84</b>

**Table 4.c :** Evaluation : Entité Nommée Caractéristiques des catastrophes naturelles

Catégorie	Précision	Rappel	F-mesure
Méthode 1	<b>0</b>	<b>0</b>	<b>0</b>
Méthode 2	<b>0.71</b>	<b>0.90</b>	<b>0.79</b>
Méthode 3	<b>0.99</b>	<b>0.95</b>	<b>0.97</b>

## **4.6 Conclusion**

Ce chapitre est la phase de réalisation de notre approche proposée, il comprend une description générale des langages et des outils utilisés pour le développement de notre application. Nous avons appliqué les règles du JAPE et les ontologies sur des documents (journaux en ligne) dans l'environnement GATE pour extraire les entités nommées Spatiale et Caractéristiques catastrophe naturelle et les dégâts, Nous avons cité les résultats sans et après l'ajout des règles JAPE et à base des ontologies.

## **Conclusion générale**

---

### **Conclusion générale**

Notre travail a pour objectif de faire l'extraction de l'information à partir de l'ontologie. Comme première étape nous nous sommes focalisés sur les thèmes de recherche qui sont en relation avec la reconnaissance des entités nommées, des ontologies, et plus précisément les thèmes qui traitent l'extraction de l'information basée ontologies.

Dans les chapitres I et II, nous avons présenté le concept général de l'extraction d'information, les entités nommées et les systèmes d'extraction basées sur les ontologies, en commençant par implanter les différents concepts permettant de cerner la notion d'entités nommée à partir d'ontologie.

Dans le chapitre III, nous avons mis l'accent sur notre système proposé qui se compose de plusieurs étapes tels que : collection des données, création d'ontologie de domaine, analyse des informations et les règles d'extraction. L'approche d'extraction d'information proposée basée sur l'ontologie à partir des documents sur le web dans le domaine de catastrophes naturelles, par la suite nous avons présenté en détaille le modèle conceptuel d'extraction d'information basé ontologie.

Dans le dernier chapitre nous avons réalisé deux approches pour l'extraction d'information et voir la différence entre elles qui sont pour objectif d'extraire les entités nommées « location », « dégâts » et « caractéristiques », dans des documents importées en ligne.

Notre principal but est l'extraction des entités nommées la plus précisément celles qui existent sur un ou plusieurs documents à l'aide d'ontologie et d'évaluer meilleure solution par rapport aux autres approches.

Pour une meilleure évaluation et optimisation de notre approche et afin de compléter ce travail et augmenter la reconnaissance des entités nommées, il est souhaitable d'utiliser différents moyens basés sur l'apprentissage automatique.



## Bibliographie

- [1] EVEN F, Extraction d'information et modélisation de connaissance à partir de Naute de communication orale , Informatique. Université de Naute, 2005, 253.
- [2] GHOULAM A, L'extraction d'information pour la recherche dans un système médical à large échelle , Informatique. Université d'Oran, 2018, 158.
- [3] TELLIR I, TAL et extraction d'information , France : Université Paris 3- Sorbonne Nouvelle, 2014, 49. (4-TELLIER)
- [4] SEGHIRI N, Détection et extraction d'information temporelles dans un entrepôts de données , Tizi-Ouzou : Université Mouloud Mammeri, 2014, 88.
- [5] : AIT RADI T et IGGUI T , Extraction d'information à base d'ontologie pour la construction d'un profil utilisateur, Mémoire de Master En Informatique, Université A.Mira Béjaïa ,2013.
- [6]SAH ,SHAGAN, Machine Learning :A Review of Learning Types,Etats-Unis :Rochester Institute of Technology, (2020), 2.
- [7] <https://larevueia.fr/apprentissage-par-renforcement/>.
- [8] DEFFAF F, Extraction des entités nommées par projection cross-linguistique et construction de lexique bilingues d'entités nommées pour la traduction automatique statique , mémoire présenté comme exigence partielle de la maitrise en informatique, Université du Québec à Montréal, 2015.
- [9] Daya C Wimalasuriya and Dejing Dou, Ontology-based information extraction: An introduction and a survey of current approaches, 2010.
- [10] WIMALASURIYA D-C, Ontology Based Information Extraction , Etats Unis : Université de l'Oregon, 2009, 37.
- [11] : VICTORRI B, MATHET Y, ENJALBERT P , Nouvelle perspective en extraction d'information , Manuscrit auteur, 2002, 19.
- [12] FLITTI S , Identification Automatique d'Entités Nommées, MEMOIRE DE FIN D'ETUDES Pour l'Obtention de Master en Informatique , MOSTAGANEM ,UNIVERSITE ABDELHAMID IBN BADIS ,2017.
- [13] DAMIEN N, Reconnaissance des entités nommées par exploration de règles d'annotation Interpréter les marqueurs d'annotation comme instructions de structuration

locale , thèse de 50 docteur en Informatique, l'École Doctorale MIPTIS, France, l'université François Rabelais de Tours, 2012.

[14] BERDANE F, HOUALI N, REZZOUG F , Modélé de langue mixte combinant entitée nommées et mots simples, Tizi-Ouzou ,Universite Mouloud Mammeri ,2013

[15] MAEDCHE A, STAAB S. The Text-To-Onto Ontology Learning Environment. Software Demonstration at the Eighth International Conference on Conceptual Structures, Berlin: Springer-Verlag; 2000.

[16] ] Yip C, Palaniappan, S. Yahaya, N.A., Health ontology system, 2011 7th International Conference on Information Technology in Asia (CITA 11), 12-13 July 2011.

[17] ALATRASH E., Using Web Tools for Constructing an Ontology of Different Natural Languages , Doctoral dissertation, University of Belgrade, 2013.

[18] <http://dspace.univ-tlemcen.dz>

[19] MAYNARD D, Peters W, Li Y. Metrics for Evaluation of Ontology-based Information. WWW 2006 Workshop, Scotland: 2006.

[20] DAVIES J, FENSL D, & Van Harmelen, F. (Eds.), Towards the semantic web: ontology-driven knowledge management , John Wiley & Sons, 2003.

[21] <https://hal.archives-ouvertes.fr/hal-00601801> Submitted on 20 Jun 2011.

[22] Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, and Angel Kirilov. "KIM—a semantic platform for information extraction and retrieval". In: Natural language engineering 10.3-4 (2004), pp. 375–392.

[23] Borislav Popov, Atanas Kiryakov, Angel Kirilov, Dimitar Manov, Damyan Ognyanoff, and Miroslav Goranov. "KIM—semantic annotation platform". In: International Semantic Web Conference. Springer. 2003, pp. 834–849.

[24] Diana Maynard, Adam Funk, and Wim Peters. "SPRAT: a tool for automatic semantic pattern-based ontology population". In: International conference for digital libraries and the semantic web. Vol. 71. 2009

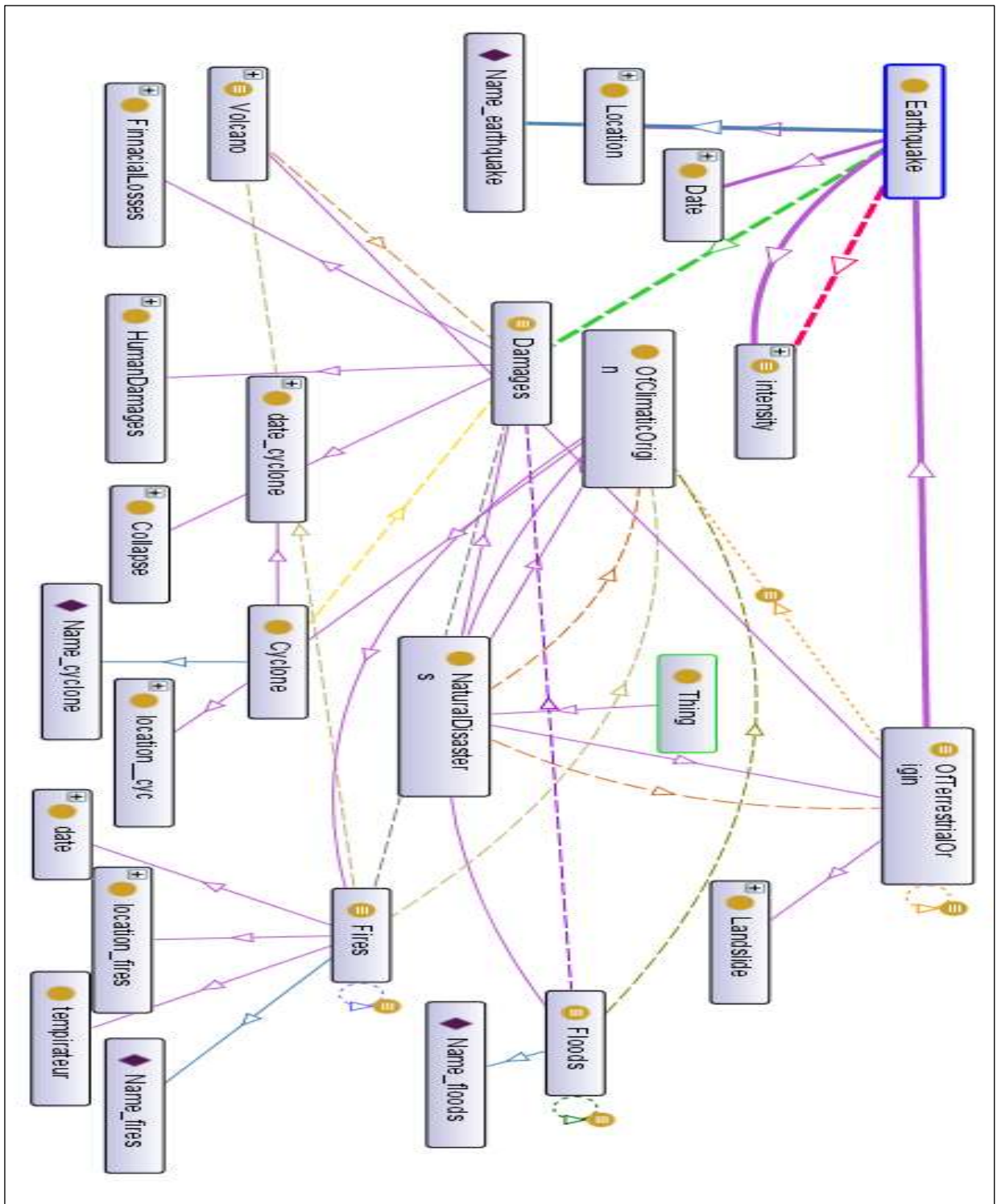
[25] Marti A Hearst. "Automatic acquisition of hyponyms from large text corpora". In: Proceedings of the 14th conference on Computational linguistics-Volume 2. Association for Computational Linguistics. 1992, pp. 539–545.

[26] Hamish Cunningham, Diana Maynard, and Valentin Tablan. "JAPE: A Java Annotation Patterns Engine, 2000". In: Research Memorandum, CS-00-10. Department of Computer Science, University of Sheffield (Dec. 1999).

- [27] Alexander Hogenboom, Frederik Hogenboom, Flavius Frasinca, Kim Schouten, and Otto Van Der Meer. “Semantics-based information extraction for detecting economic events”. In: *Multimedia Tools and Applications* 64.1 (2013), pp. 27–52.
- [28] Roberto Navigli and Paola Velardi. “Structural semantic interconnections: a knowledge-based approach to word sense disambiguation”. In: *IEEE transactions on pattern analysis and machine intelligence* 27.7 (2005), pp. 1075–1086.
- [29] <https://gate.ac.uk/sale/tao/splitch6.html#x9-1210006.2>
- [30] Protege Software, <http://protege.stanford.edu/>, 2015, January, 15.
- [31] Bhaskar, K., Savita, S., “A Comparative Study of Ontology building Tools in Semantic Web Applications”, *International journal of Web & Semantic Technology (IJWesT)* Vol.1, No.3, July 2010.
- [32] ] GATE Software, <https://gate.ac.uk/>, 2014, November, 18.
- [33] Ranganathan, G.; Biletskiy, Y.; Kaltchenko, A., “Semantic annotation of semi-structured documents”, *Canadian Conference on Electrical and Computer Engineering*, 2008. CCECE 2008, pp. 919, 922, 4-7 May 2008.

Annexes:

Annexe 1 : Ontologie des catastrophes naturelles



## Annexe 2: Une partie du code source de l'ontologie de NDO dans Owl

```
es.html x3 OND.owl x3 OND.owl x3
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#"
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#"
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#"
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
]>

<rdf:RDF xmlns="http://www.semanticweb.org/pc/ontologies/2022/4/OND#"
  xml:base="http://www.semanticweb.org/pc/ontologies/2022/4/OND"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <owl:Ontology rdf:about="http://www.semanticweb.org/pc/ontologies/2022/4/OND">
    <rdfs:comment rdf:datatype="&xsd:string">The ontology about disester Naturelle
    </rdfs:comment>
  </owl:Ontology>
```

```
<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity -->
<owl:Class rdf:about="http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity">
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity\_strength" />
      <owl:onClass rdf:resource="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity" />
      <owl:minQualifiedCardinality rdf:datatype="&xsd:nonNegativeInteger">2
      </owl:minQualifiedCardinality>
    </owl:Restriction>
  </owl:equivalentClass>
  <owl:equivalentClass>
    <owl:Restriction>
      <owl:onProperty rdf:resource="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity\_strength" />
      <owl:onClass rdf:resource="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#intensity" />
      <owl:maxQualifiedCardinality rdf:datatype="&xsd:nonNegativeInteger">9
      </owl:maxQualifiedCardinality>
    </owl:Restriction>
  </owl:equivalentClass>
  <rdfs:subClassOf rdf:resource="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#Earthquake" />
</owl:Class>
```



```
<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#Has\_Earthquake -->
```

```
<owl:ObjectProperty rdf:about="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#Has\_Earthquake">
  <rdfs:range rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#Damages" />
  <rdfs:domain rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#Earthquake" />
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#High\_temperateure -->
```

```
<owl:ObjectProperty rdf:about="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#High\_temperateure">
  <rdfs:domain rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#Fires" />
  <rdfs:range rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#OfClimaticOrigin" />
  <rdfs:range rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#Volcano" />
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#Rainfall -->
```

```
<owl:ObjectProperty rdf:about="
http://www.semanticweb.org/pc/ontologies/2022/4/OND#Rainfall">
  <rdfs:domain rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#Floods" />
  <rdfs:range rdf:resource="
    http://www.semanticweb.org/pc/ontologies/2022/4/OND#OfClimaticOrigin" />
```

```
<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#schoolles -->

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/pc/ontologies/2022/4/OND#schoolles">
  <rdf:type rdf:resource="http://www.semanticweb.org/pc/ontologies/2022/4/OND#Collapse"/>
</owl:NamedIndividual>

<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#the\_missing -->

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/pc/ontologies/2022/4/OND#the_missing">
  <rdf:type rdf:resource="http://www.semanticweb.org/pc/ontologies/2022/4/OND#HumanDamages"/>
</owl:NamedIndividual>

<!-- http://www.semanticweb.org/pc/ontologies/2022/4/OND#wounded -->

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/pc/ontologies/2022/4/OND#wounded">
  <rdf:type rdf:resource="http://www.semanticweb.org/pc/ontologies/2022/4/OND#HumanDamages"/>
</owl:NamedIndividual>
</rdf:RDF>

<!-- Generated by the OWL API (version 3.4.2) http://owlapi.sourceforge.net -->
```